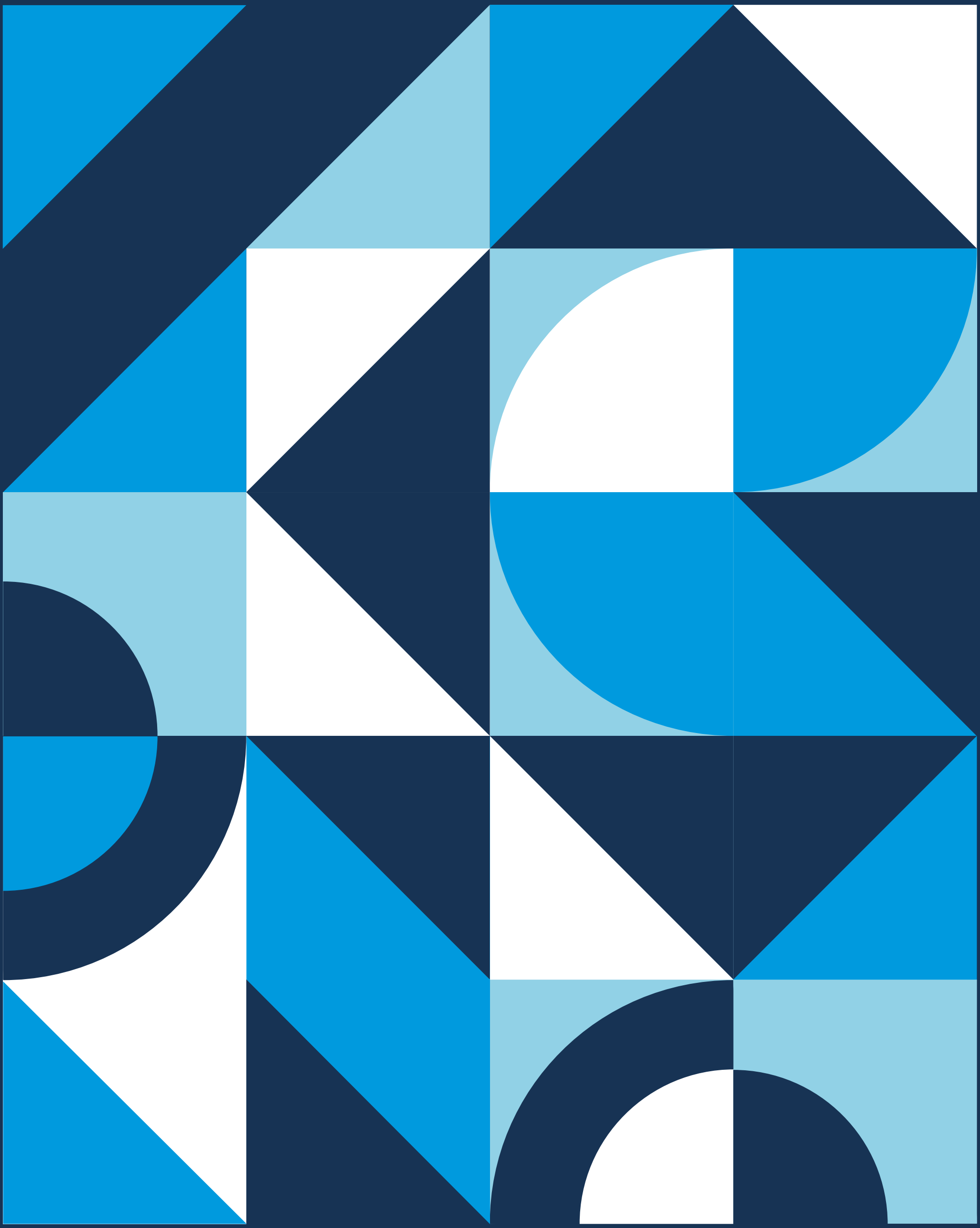
A MANIFESTO

for Applying Behavioral Science

THE
BEHAVIOURAL
INSIGHTS
TEAM



Michael Hallsworth

ACKNOWLEDGMENTS

Thanks to Lila Tublin for drafting and editing support, Richard O'Brien for communications support, Dilhan Perera for thoughts on the "Predict and Adjust" section, and Alex Gyani for his input on the "Replication, Variation, Adaptation" section.

Thanks to S. Banerjee, E. Berkman, A. Buttenheim, F. Callaway, J. Collins, E. Costa, J. Doctor, D. Halpern, P. John, S. Kousta, T. Marteau, M. Muthukrishna, C. Payne, D. Perrott, K. Ruggeri, R. Schmidt, D. Soman, H. Strassheim, C. Sunstein, and members of the Behavioural Insights Team for their feedback on previous drafts. I am grateful to the four peer reviewers at Nature Human Behaviour for their helpful suggestions.

SUMMARY

This is a manifesto for how applied behavioral science can fulfill its true potential.

The behavioral insights movement has flourished over the last decade.[1] There is now a vibrant ecosystem of practitioners, teams, and academics building on each other's work across the globe. Their focus on robust evaluation means we know that this work has had an impact on important issues such as antimicrobial resistance, educational attainment, climate change, and obesity.

The Behavioural Insights Team is proud to have been a pioneer of this growth. However, we and others in the field also realize that behavioral science needs to evolve further over its next decade.

In this manifesto we take a clear-eyed look at the challenges facing the field and offer ten proposals for making further progress. As a starting point, we present the main arguments from critics of the behavioral insights approach on the following page.

THE CRITICISMS

Limited impact

Limited impact. The approach has focused on more tractable and easy-to-measure changes at the expense of bigger impact: it has just been tinkering around the edges of fundamental problems.[2]

Failure to reach scale

The approach promotes a model of experimentation followed by scaling, but it has not paid enough attention to how successful scaling happens - and the fact it often does not happen.[3]

Mechanistic thinking

The approach has promoted a simple, linear, and mechanistic way of understanding and influencing behavior that ignores second-order effects and spillovers (and employs evaluation methods that assume a move from A to B against a static background).[4]

Flawed evidence base

The replication crisis has challenged the evidence base underpinning the behavioral insights approach, adding to existing concerns like the duration of its interventions' effects.[5]

Lack of precision

The approach lacks the ability to construct precise interventions and establish what works for whom, and when. Instead, it relies either on over-general frameworks or disconnected lists of biases.[6]

Overconfidence

The approach is affected by the wider problem of over-confidence and can over-extrapolate from its evidence base, particularly when testing is not an option.[7]

Control paradigm

The approach can be elitist and pays insufficient attention to people's own goals and strategies; it uses concepts like "irrationality" to justify attempts to control the behavior of individuals, since they lack the means to do so themselves.[8]

Neglect of the social context

The approach has a limited, overly cognitive and individualistic view of behavior that neglects the reality that humans are embedded in established societies and practices.[9]

Ethical concerns

The behavioral insights approach will face more ethics, transparency, and privacy conundrums as it attempts more ambitious and innovative work.[10]

Homogeneity of participants and perspectives

The range of participants in behavioral science research has been narrow and unrepresentative; [11] homogeneity in the locations and personal characteristics of behavioral scientists influences their viewpoints, practices, and theories.[12]

THE PROPOSALS

We do not agree with all these criticisms, but we do think that they highlight several challenges that must - and can - be met. Doing so will mean behavioral science is better equipped to help build policies, products, and services on stronger empirical foundations - and thereby address the world's crucial challenges.

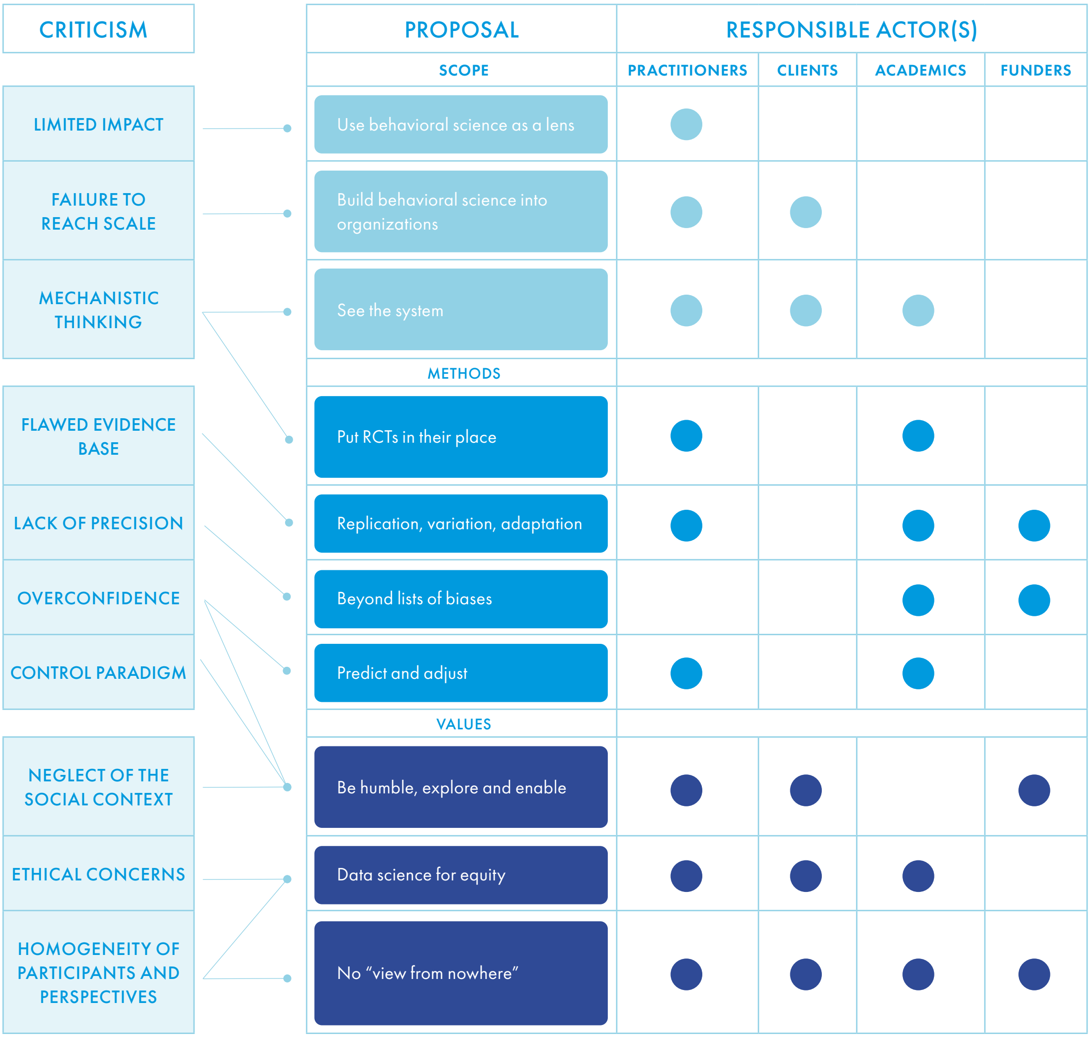
Our ten proposals for applied behavioral science fall into three categories: scope (the range and scale of issues to which behavioral science is applied); methods (the techniques and resources that behavioral science deploys); and values (the principles, ideals, and standards of conduct that behavioral scientists adopt).

Category	Proposal	Recommended action(s)
Scope	01 USE BEHAVIORAL SCIENCE AS A LENS Summary: Page 11 Detail: Page 25	Present behavioral science as a lens that improves the view of any public and private issue, in order to break a self-sustaining pattern that has directed behavioral science away from the most significant problems.
	02 BUILD BEHAVIORAL SCIENCE INTO ORGANIZATIONS Summary: Page 12 Detail: Page 29	Focus less on how to set up a dedicated behavioral science team, and more on how the approach can be integrated into an organization's standard processes by upgrading its "choice infrastructure".
	03 SEE THE SYSTEM Summary: Page 14 Detail: Page 37	Use aspects of complexity thinking to improve behavioral science so it can: exploit "leverage points"; model the collective implications of heuristics; alter specific features of systems to create wider changes; and understand the longer-term impact on a system of a collection of policies with varying goals.

Category	Proposal	Recommended action(s)
Methods	**04** **PUT RCTs IN THEIR PLACE** Summary: Page 15 Detail: Page 45	Strengthen RCTs to deal better with complexity by: gaining a better understanding of the system interactions and anticipate how they may play out; setting up RCTs to measure diffusion and contagion in networks; building feedback and adaptation into the design of RCTs and interventions.
	05 **REPLICATION, VARIATION AND ADAPTATION** Summary: Page 16 Detail: Page 51	Identify the most reliable interventions, develop an accurate sense of the likely size of their effects, and avoid the weaker options. Recognize that heterogeneity requires a much higher bar for claiming that an effect holds true across many unspecified settings. Create multi-site studies to systematically study heterogeneity in a wider range of contexts and participants. Codify and cultivate the practical skills that successfully adapt interventions to new contexts.
	06 **BEYOND LISTS OF BIASES** Summary: Page 17 Detail: Page 61	Emphasize theories that are "practical": they fill the gap between high-level frameworks and jumbled lists of biases; they are based on data and generate testable hypotheses, but also specify the conditions under which a prediction applies; they present actionable steps to solve real-world problems.
	07 **PREDICT AND ADJUST** Summary: Page 18 Detail: Page 69	Develop the practice of getting behavioral scientists to predict the results of experiments, and then feeding back the results to them.

Category	Proposal	Recommended action(s)
Values	08 **BE HUMBLE, EXPLORE AND ENABLE** Summary: Page 18 Detail: Page 75	Avoid using the term "irrationality"; practice "epistemic humility"; and design processes and institutions to counteract overconfidence. Pay greater attention to people's own interpretations of their beliefs, feelings and behaviors. Reach a wider range of experiences, including marginalized voices and communities. Recognize how apparently universal cognitive processes are shaped by specific contexts. Use six criteria (detailed in the main text) to assess when to enable people to use behavioral science themselves.
	09 **DATA SCIENCE FOR EQUITY** Summary: Page 20 Detail: Page 85	Use data science to identify the ways in which an intervention or situation appears to increase inequalities and introduce features to reduce them. For example, groups that are particularly likely to miss a filing requirement could be offered pre-emptive help.
	10 **NO VIEW FROM NOWHERE** Summary: Page 21 Detail: Page 91	Cultivate self-scrutiny; find new ways for the subjects of research to judge researchers; take actions to increase diversity among behavioral scientists and their teams, such as building professional networks between the Global North and Global South.

The figure below shows how each proposal maps onto the criticisms, as well as which groups have responsibility for implementing them: practitioners (individuals or teams who apply behavioral science findings in practical settings); the clients who commission these practitioners (for example, public or private sector organizations); academics working in the behavioral sciences (including disciplines such as anthropology, economics, and sociology); and funders who support the work of these academics.



CRITICISM

LIMITED IMPACT

FAILURE TO REACH SCALE

MECHANISTIC THINKING

FLAWED EVIDENCE BASE

LACK OF PRECISION

OVERCONFIDENCE

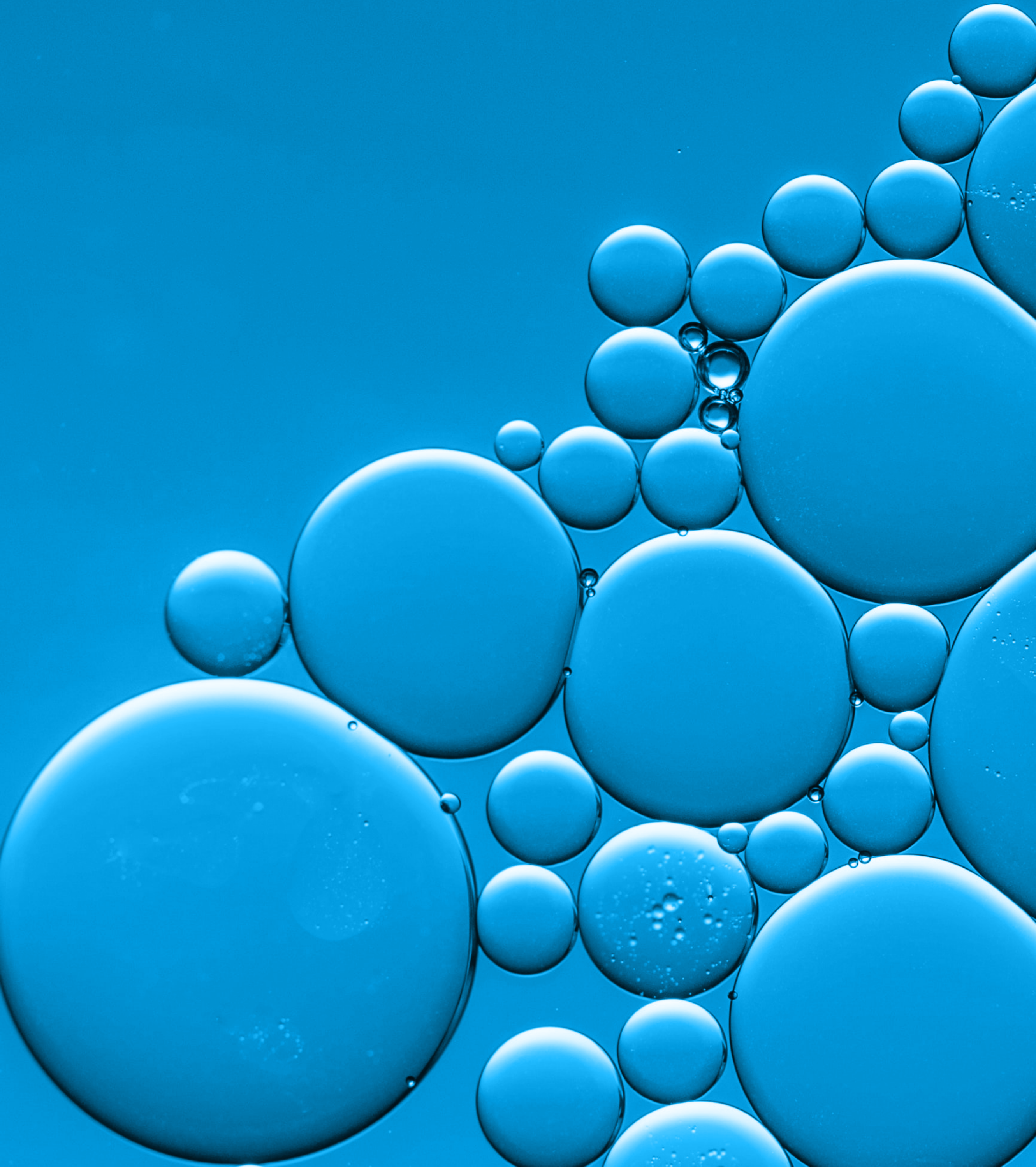
CONTROL PARADIGM

NEGLECT OF THE SOCIAL CONTEXT

ETHICAL CONCERNS

HOMOGENEITY OF PARTICIPANTS AND PERSPECTIVES

PROPOSAL	RESPONSIBLE ACTOR(S)			
SCOPE	PRACTITIONERS	CLIENTS	ACADEMICS	FUNDERS
Use behavioral science as a lens	●			
Build behavioral science into organizations	●	●		
See the system	●	●	●	
METHODS				
Put RCTs in their place	●		●	
Replication, variation, adaptation	●		●	●
Beyond lists of biases			●	●
Predict and adjust	●		●	
VALUES				
Be humble, explore and enable	●	●		●
Data science for equity	●	●	●	
No "view from nowhere"	●	●	●	●



SCOPE

01

USE BEHAVIORAL SCIENCE AS A LENS

The early phase of the behavioral insights movement was marked by skepticism about whether findings from laboratories would translate to real-world settings.[13] Mindful of this concern, practitioners developed standard approaches that could demonstrate a clear causal link between an intervention and an outcome.[14]

In practice, these approaches directed attention towards how the design of specific aspects of a policy, product or service influences discrete behaviors by actors who are considered mostly in isolation.[15] These standard approaches are strong and have produced compelling results. But they have also encouraged people to see behavioral science as a kind of specialist tool. This view mostly limits behavioral science to fixing concrete aspects of predetermined interventions - rather than shaping broader policy goals. Behavioral science acts as an alternative to standard tools, and it should be applied only to certain kinds of "behavioral" issues.[16]

Such a view is both misguided and profoundly limiting, but over time it has created a self-reinforcing perception that only certain kinds of tasks are "suitable" for behavioral scientists.[17] Opportunities, skills and ambitions have been constricted as a result.

A rebalancing is needed. Behavioral science also has much to say about pressing societal issues like discrimination, pollution, or economic mobility, and the structures that produce them.[18] These ambitions have always been present in the behavioral insights movement,[19] but the factors just outlined acted against them being realized more fully.[20]

The first step is to change the way we frame behavioral science itself. We need to see behavioral science as a lens that can be applied to any public and private issue. Using this frame shows that behavioral insights can enhance the way we see policy options (for example, revealing new ways of structuring taxes), rather than just acting as an alternative to them; it also conveys that creating new interventions to change behavior is not always the goal - which means more weight should be placed on the behavioral diagnosis of an issue. Behavioral science itself shows us the power of framing: the metaphors we use shape the way we behave, and therefore can be an agent of change.[21] Metaphors are particularly important in this case because the task of broadening the use of behavioral science requires making a compelling case to decision makers.[22] The metaphor of behavioral insights as a tool has established credibility and acceptance in a defined area; expanding beyond that area is the task for the next decade.

Full detail on page 25

SCOPE

02 BUILD BEHAVIORAL SCIENCE INTO ORGANIZATIONS

There has been too little focus on using behavioral science to shape organizations themselves, as opposed to increasing how much an organization uses behavioral science to achieve its goals.[23] We need to talk less on how to set up a dedicated behavioral team, and more about how behavioral science can be integrated into an organization's standard processes. For example, as well as trying to ensure that a departmental budget includes provisions for behavioral science, why not use behavioral science to improve the way this budget is created (e.g., are managers anchored to outdated spending assumptions)?

But we need to understand how this new way of thinking maps against existing debates about how to set up a behavioral function in organizations. We propose that doing so reveals six main scenarios, as shown in the diagram below.

In the "Baseline" scenario there is limited awareness of behavioral science in the organization, and its principles are not incorporated into processes. In the "Nudged Organization," levels of behavioral science awareness are still low, but its principles have been used to redesign processes to create better outcomes for staff or service users. No explicit behavioral science knowledge or capacity is created or needed, which means the return on investment here could be large. For that reason, this model feels like a neglected opportunity.

In "Proactive Consultancy", leaders may have set up a dedicated behavioral team without enough supporting organizational changes. The result is that the team has to work in an enterprising way, going to look for opportunities and having to prove its worth. But these teams may not be in a resilient position, since they lack ways to be grafted onto the standard processes of an organization.

Greater potential for scale

		BEHAVIORAL SCIENCE KNOWLEDGE AND CAPACITY		
		LIMITED	CONCENTRATED	DIFFUSED
BEHAVIORAL SCIENCE INCORPORATED INTO ORGANIZATIONAL PROCESSES	NO	Baseline	Proactive consultancy	Behavioral entrepreneurs
	YES	Nudged organization	"Call for the experts"	Behaviorally-enabled organization

In "Call For The Experts", an organization has concentrated behavioral expertise, but there are also prompts and resources that allow this expertise to be integrated more into "business as usual". Expertise is not widespread, but access to it is. This setup could mean that processes stimulate demand for behavioral expertise that the central team can fulfill. That team may also have the institutional support to proactively monitor activities and respond quickly to specific crises.

In "Behavioral Entrepreneurs", there is behavioral science capacity distributed throughout the organization, either through direct capacity building or recruitment. The problem is that organizational processes do not support these individual pockets of knowledge. Therefore, those with expertise find it hard to apply ideas in practice, evaluate their effects, share findings, and build learning.

Finally, a "Behaviorally-Enabled Organization" is one where there is knowledge of behavioral science diffused throughout the organization, which also has processes that reflect this knowledge and support its deployment. Staff apply behavioral science in a deliberate way as part of "business as usual", rather than as special projects. While this is the most resilient setup, it also requires the most resources.

Most discussions make it seem like the meaningful choice is between the different columns in the table above - how to organize dedicated behavioral science resources. Instead, the more important move is from the top row to the bottom row: moving from projects to processes, from commissions to culture. A useful way of thinking about this task is about building or upgrading the "choice infrastructure" of the organization.[24]

Working out how best to build the choice infrastructure in organizations should be a major priority for behavioral science. One advantage to this approach is that it can help organizations address problems with scaling interventions. Already we can see some features will be crucial: reducing the costs of experimentation; creating a system that can learn from its actions; and developing new and better ways of using behavioral science principles to analyze the behavioral effects of organizational processes, rules, incentives, metrics, and guidelines.[25]

Full detail on page 29

SCOPE

03

SEE THE SYSTEM

Many important policy challenges emerge from complex adaptive systems, where change often does not happen in a linear or easily predictable way, and where coherent behavior can emerge from interactions without top-down direction.[26] There are many examples of such systems in human societies, including cities, markets, and political movements.[27] These systems can create “wicked problems” - like the Covid-19 pandemic - where ideas of success are contested, changes are non-linear and difficult to model, and policies have unintended consequences.[28]

This reality challenges the dominant behavioral science approach, which usually assumes stability over time, keeps a tight focus on predefined target behaviors, and predicts linear effects based on a predetermined theory of change.[29] The end result, some argue, is a failure to understand how actors are acting and reacting in a complex system that leads policymakers to conclude they are being “irrational” - and then actually disrupt the system in misguided attempts to correct perceived biases.[30] Behavioral science can be improved by using aspects of complexity thinking to offer new, credible, and practical ways of addressing major policy issues. First, we need to reject crude distinctions of “upstream” versus “downstream” or the “individual frame” versus the “system frame”.[31] Instead, complex adaptive systems show that “higher-level” features of a system can actually emerge from the “lower-level” interactions of actors participating in the system.[32] When they become the governing features of the system, they then shape the “lower-level” behavior until some other aspect emerges, and the fluctuations continue. We can see this pattern in the way that new coronavirus variants emerged from specific contexts to re-shape the whole course of the pandemic.

In other words, we are dealing with ‘cross-scale behaviors’.[33] For example, norms, rules, practices, and culture itself can emerge from aggregated social interactions; these features then shape cognition and behavioral patterns in turn.[34]

Recognizing cross-scale behaviors means that behavioral science could:

- Identify “leverage points’’ where a specific shift in behavior will produce wider system effects. For example, if even a subset of consumers decides to switch to a healthier version of a food product, this can have broader effects on a population’s consumption through the way the food system responds by restocking and product reformulation.[35]
- Model the collective implications of individuals using simple heuristics to navigate a system. For example, new models show how small changes to simple heuristics that guide savings (in this case, how quickly households copy the savings behaviors of neighbors) can lead to inequalities in wealth suddenly emerging.[36]
- Find targeted changes to features of a system that create the conditions for wide-ranging shifts in behavior to occur. For example, a core driver for social media behaviors is the ease with which information can be shared.[37] Even minor changes to this factor can drive widespread changes - some have argued that such a change is what created the conditions leading to the Arab Spring, for example.[38]

This approach also suggests that a broader change in perspective is needed. We need to realize the flaws in launching interventions in isolation and then moving on when a narrowly defined goal has been achieved. Instead, we need to see the longer-term impact on a system of a collection of policies with varying goals.[39] The best approach may be “system stewardship”, which focuses on creating the conditions for behaviors and indirectly steering adaptation towards overall goals.[40]

Of course, not every problem will involve a complex adaptive system; for simple issues the standard behavioral approach works well. So behavioral scientists should develop the skills to recognize the type of system that they are facing (“see the system”), and then choose their approach accordingly. These skills can be developed through agent-based simulations,[41] immersive technologies,[42] or just basic checklists.[43]

Full detail on page 37

METHODS

04

PUT RCTs IN THEIR PLACE

Randomized Controlled Trials (RCTs) have been a core part of applied behavioral science, and they work very well in relatively simple and stable contexts. But they can fare worse in complex adaptive systems, whose many shifting connections can make it difficult to keep a control group isolated, and where a narrow focus on predetermined outcomes may neglect others that are important but difficult to predict.[44]

We can strengthen RCTs to deal better with complexity. We can try to gain a better understanding of the system interactions and anticipate how they may play out, perhaps through "dark logic" exercises that try to trace potential harms, rather than benefits.[45] We can set up RCTs to measure diffusion and contagion in networks, either by creating separate online environments or by randomizing real-world clusters, like separate villages.[46]

Finally, we can build feedback and adaptation into the RCT design, allowing adjustments to changing conditions.[47] Options include using two-stage trial protocols,[48] evolutionary RCTs,[49] sequential multiple assignment randomized (SMART) trials,[50] and "bandit" algorithms that identify high-performing interventions and allocate more people to them.[51]

We can also use behavioral science to enhance alternative ways of measuring impact - in particular, agent-based modeling, which tries to simulate the interactions between the different actors in a system.[52] The agents in these models are mostly assumed to be operating on rational choice principles.[53] Therefore, there is a big opportunity to build in more evidence about the drivers of behavior - for example, habits and social comparisons.[54]

Full detail on page 45

METHODS

05

REPLICATION, VARIATION, ADAPTATION

The "replication crisis" of the last decade has seen intense debate and concern about the reliability of behavioral science findings. Poor research practices were a major cause of the replication crisis; the good news is that many have improved as a result.[55]

We need to secure and build on these advances, so that we move towards a future where meta-analyses of high-quality studies (including deliberate replications) are used to identify the most reliable interventions, develop an accurate sense of the likely size of their effects, and avoid the weaker options. We have a responsibility to discard ideas if solid evidence now shows they are shaky, and to offer a realistic view of what behavioral science can accomplish.

That responsibility also requires us to have a hard conversation about heterogeneity in results: the complexity of human behavior creates so much statistical "noise" that it's often hard to detect consistent signals and patterns.[56] The main drivers of heterogeneity are that a) contexts influence results and b) the effect of an intervention may vary greatly between groups within a population.[57] These factors complicate the idea of replication itself: a "failed" replication may not show that a finding was false, but rather how it exists under some conditions and not others.[58]

These challenges mean that applied behavioral scientists need to set a much higher bar for claiming that an effect holds true across many unspecified settings.[59] There is a growing sense that interventions should be talked about as hypotheses that were true in one place, and which may need adapting for them to be true elsewhere as well.[60]

We need specific proposals as well as narrative changes. The first concerns data collection: expand studies to include (and thus examine) a wider range of contexts and participants, and gather richer data about them. To date, only a small minority of behavioral studies have provided enough information to see how effects vary.[61] Coordinated multi-site studies will be needed to collect enough data to explore heterogeneity systematically; "crowdsourced" studies offer particular promise for testing context and methods.[62]

Behavioral scientists also need to get better at judging how much an intervention's results were linked to its context - and therefore how much adaptation it may need.[63] We should use and modify frameworks from implementation science to develop such judgment.[64] Finally, we need to codify and cultivate the practical skills that successfully adapt interventions to new contexts; expertise in behavioral science should not be seen as simply knowing about concepts and findings in the abstract. Therefore, it's particularly valuable to learn from practitioners how they adapted specific interventions to new contexts. These accounts are starting to emerge, but they are still rare,[65] since researchers are incentivized to claim universality for their results, rather than report and value contextual details.[66]

Full detail on page 51

METHODS

06 BEYOND LISTS OF BIASES

The heterogeneity in behavioral science findings also means that our underlying theories need to improve: we are lacking good explanations for why findings vary so much.[67] This need for better theories can be seen as part of a wider "theory crisis" in psychology, which has thrown up two big concerns for behavioral science.[68]

The first stems from the fact that theories of behaviour often try to explain phenomena that are complex and wide-ranging.[69] Trying to cover this variability can produce descriptions of relationships and definitions of constructs that are abstract and imprecise. The result is theories that are vague and "weak", since they can be used to generate many different hypotheses - some of which may actually contradict each other.[70] That makes theories hard to disprove and so weak theories stumble on, unimproved.[71]

The other concern is that theories can make specific predictions, but they are disconnected from each other - and from a deeper, general framework that can provide broader explanations (like evolutionary theory, for example).[72] The main way this issue affects behavioral science is through heuristics and biases. Examples of individual biases are accessible, popular, and how many people first encounter behavioral science. These ideas are incredibly useful, but have often been presented as lists of standalone curiosities, in a way that is incoherent, reductive, and deadening. They can create overconfident thinking that targeting a specific bias (in isolation) will achieve a certain outcome.[73]

Perhaps most importantly, focusing on lists of biases distracts us from answering core underlying questions. When does one or another bias apply? Which are widely applicable, and which are highly specific? These are highly practical questions when someone is faced with tasks like, for example, taking an intervention to new places.

The concern for behavioral science is that it uses both these high-level frameworks, like dual process theories, and jumbled collections of heuristics and biases - with little in the middle to draw both levels together.[74]

We think that a priority for responding to this challenge is to develop theories that are practical. By this we mean:

- They fill the gap we've identified in behavioral science: between day-to-day working hypotheses and comprehensive and systematic attempts to find universal underlying explanations.
- They are based on data rather than being derived from pure theorizing.[75]
- They can generate testable hypotheses, so they can be disproved.[76]
- However, they also specify the conditions under which a prediction applies or does not.[77]
- They are geared towards realistic adaptation by practitioners and offer 'actionable steps toward solving a problem that currently exists in a particular context in the real world.'[78]

We think that resource rationality is a good example of a practical theory. It starts from the basis that people make rational use of their limited cognitive resources.[79] Given there is a cost to thinking, people will look for solutions that balance choice quality with effort. Importantly, these principles offer a systematic framework for building useful models for how people act.

A recent study has shown how these models can can not only predict how people will respond to different kinds of nudges in certain contexts, but also can be integrated with machine learning to create an automated method for constructing 'optimal nudges'.[80] These are highly practical benefits coming from applying a particular theory.

Full detail on page 61

METHODS

07 PREDICT AND ADJUST

Hindsight bias is what happens when people feel "I knew it all along", even if they did not.[81] When the results of an experiment come in, hindsight bias may mean that behavioral scientists are more likely to think that they had predicted them, or quickly find ways of explaining why they occurred. Hindsight bias is a big problem because it breeds overconfidence, impedes learning, dissuades innovation, and prevents us from understanding what is truly unexpected.[82]

In response, behavioral scientists should establish a standard practice of predicting the results of experiments, and then receiving feedback on how their predictions performed. Hindsight bias can flourish if we do not systematically capture expectations or "priors" about what the results of a study will be - in other words, it is not easy to check or remember the state of knowledge before an experiment.[83] Making predictions provides regular, clear feedback of the kind that is more likely to trigger surprise and reassessment, rather than hindsight bias.[84]

More and more studies are explicitly integrating predictions.[85] But barriers lie in the way of further progress. People may not welcome the ensuing challenge to their self-image; predicting may seem like one thing too many on the to-do list; and the benefits lie in the future.

We propose: make predicting easy by incorporating it into standard organizational processes; minimize threats to predictors' self-image, for example by making and feeding back predictions anonymously;[86] give concrete prompts for learning and reflection, in order to disrupt the move from surprise to hindsight bias;[87] and build learning from prediction within and between institutions.

Full detail on page 69

VALUES

08 BE HUMBLE, EXPLORE AND ENABLE

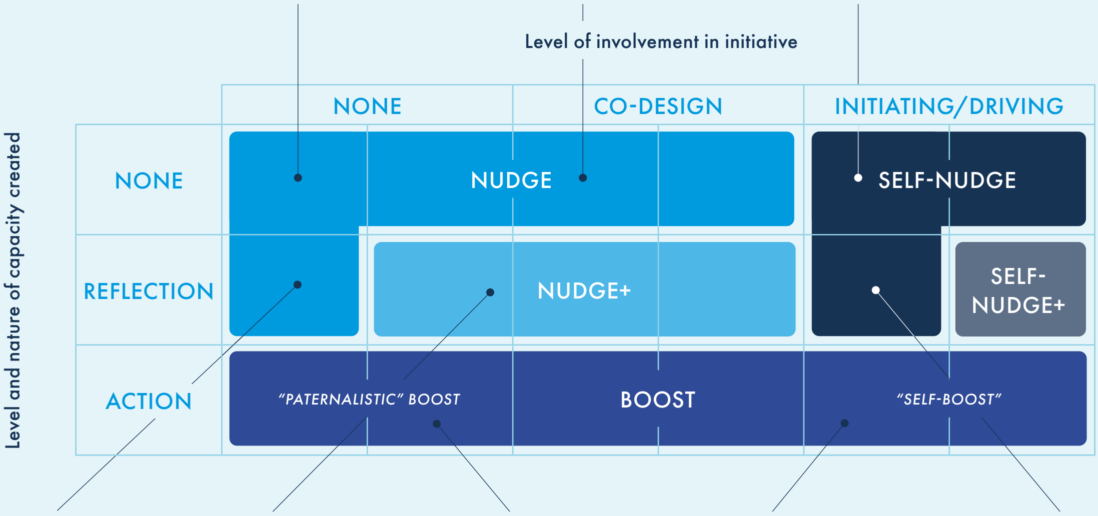
Behavioral scientists (like other experts) may over-confidently rely on decontextualized principles that do not match the real-world setting for a behavior.[88] Deeper inquiry can reveal reasonable explanations for what seem to be behavioral biases.[89] In response, those applying behavioral science should: avoid using the term "irrationality", which can limit attempts to understand actions in context; acknowledge that diagnoses of behavior are provisional and incomplete ("epistemic humility");[90] and design processes and institutions to counteract overconfidence.[91]

A common theme through these ideas is the need for more and better inquiry into behaviors in context, rather than making assumptions. Open-ended qualitative exploration of the context and drivers for behaviors is not new to the behavioral sciences.[92] However, three areas demand particular focus in the future. First, pay greater attention to people's goals and strategies, and their own interpretations of their beliefs, feelings, and behaviors.[93]
Second, reach a wider range of experiences, including marginalized voices and communities, understanding how structural inequalities can lead to expectations and experiences varying greatly by group and geography.[94] Third, recognize how apparently universal cognitive processes are shaped by specific contexts, thereby unlocking new ways for behavioral science to engage with values and culture.[95]

e.g. healthy foods placed prominently in canteen

e.g. employees deciding they want healthy foods to be prominent & developing arrangement

e.g. creating regular automatic bank transfers to savings account



e.g. gyms introducing commitment devices to increase usage

e.g. users working with gyms to develop commitment devices that actively remind users of their exercise goals

e.g. initiator decides that the best option is to teach people heuristics to increase savings

e.g. individual reads about heuristics and constructs own heuristics for savings goals

e.g. placing letters to be posted on doorhandle

In addition, more can and should be done to broaden ownership of behavioral science approaches. Many, but far from all, behavioral science applications have been quite top-down, with a 'choice architect' enabling certain outcomes.[96] One route is to enable people to become more involved in designing interventions themselves - and "nudge plus", "self nudges", and "boosts" have been proposed as ways of doing this.[97] Reliable criteria are needed to decide when enabling approaches may be appropriate, including: whether the opportunity to use an enabling approach exists; ability and motivation; preferences; learning and setup costs; equity impacts; and effectiveness (recognizing evidence on this point is still emerging).[98]

But these new approaches should not be seen simplistically as "enabling" alternatives to "disempowering" nudges.[99] Instead, we need to consider a) how far the person performing the behavior is involved in shaping the initiative itself; b) the level and nature of any capacity created by the intervention.

People may be heavily engaged in selecting and developing a nudge intervention that nonetheless does not trigger any reflection or build any skills. Alternatively, a policy maker may have paternalistically assumed that people want to build up their capacity to perform an action, when in fact they do not. This is the real choice to be made.

A final piece missing from current thinking is that enabling people can lead to a major decentering of the use of behavioral science. If more people are enabled to use behavioral science, they may decide to introduce interventions that influence others. Rather than just creating self-nudges through altering their immediate environments, they may decide that wider system changes are needed instead. A range of people could be enabled to create nudges that generate positive societal change (with no "central" actors involved), as happened for the "Fair Tax Mark" in the UK.

Full detail on page 75

VALUES

09 DATA SCIENCE FOR EQUITY

Recent years have seen growing interest in using new data science techniques to reliably analyze the heterogeneity of large datasets.[100] Machine learning is claimed to offer more sophisticated, reliable, and data-driven ways of detecting meaningful patterns in datasets.[101] For example, a machine learning approach has been shown to be more effective than conventional segmentation approaches at analyzing patterns of US household energy usage to reduce peak consumption.[102]

A popular idea is to use such techniques to better understand what works best for certain groups, and thereby tailor an offering to them.[103] "Scaling" an intervention stops being about a uniform roll-out, and instead becomes about presenting recipients with the aspects that are most effective for them.[104]

This vision is often presented as straightforward and obviously desirable, but it runs almost immediately into ethical quandaries and value judgements. People are unlikely to know what data has been used to target them, and how; the specificity of the data involved may make manipulation more likely, since it may exploit sensitive personal vulnerabilities; and expectations of universality and non-discrimination in public services may be violated.[105]

There is also emerging evidence that people often object to personalization. While they support some personalized services, they consistently oppose advertising that is customized based on sensitive information - and they are generally against the collection of the information that personalization relies on.[106] When a company tries personalization that crosses into being "creepy," uproar and damage to its reputation can ensue.[107]

In order to navigate this landscape, behavioral scientists need to examine four factors.

- **Who** does the personalization target, and using what criteria? Many places have laws or norms to ensure equal treatment based on personal characteristics. When does personalization violate those principles?
- **How** is the intervention constructed? To what extent do the recipients have awareness of the personalization, choice over whether it occurs, control over its level or nature, and the opportunity for giving feedback on it?[108]
- **When** is it directed? Is it at a time when the participant is vulnerable? Would they likely regret it later, if they had time to reflect?
- **Why** is personalization happening? Does it aim to exploit and harm or support and protect, recognizing that those terms are often contested?

Taking these factors into account, we propose that the main opportunity is for data science to identify the ways in which an intervention or situation appears to increase inequalities, and reduce them.[109] For example, groups that are particularly likely to, say, miss a filing requirement, could be offered preemptive help.

We call this idea data science for equity. It addresses the "why" factor by using data science to support not exploit. But it needs to be supported by other attempts to increase agency (the "how" factors), like a recent study that showed how boosts can be used to help people detect micro-targeting of advertising,[110] and studies that obtain more data on which uses of personalization people find acceptable.

Full detail on page 85

VALUES

10 NO VIEW FROM NOWHERE

Behavioral scientists need to understand how they bring certain assumptions, privileges, and ways of seeing to what they do.[111] They are always situated, embedded, and entangled with ideas and situations. They cannot assume there is some set-aside position from which to observe the behavior of others - there is no "view from nowhere".[112]

Behavioral scientists are defined by having knowledge, skills, and education; many of them can use these resources to shape public and private actions. Therefore, they are in a privileged position, but may not see the extent to which they hold elite positions that stop them from understanding people who think differently (for example those who are skeptical of education).[113]

There have been repeated concerns that the field is still highly homogeneous in other ways as well. Gender, race, physical abilities, sexuality, and geography also influence the viewpoints, practices, and theories of behavioral scientists.[114] Only a quarter of the behavioural insights teams catalogued in a 2020 survey were based in the Global South.[115] The last decade has shown just how behaviors can vary greatly from culture to culture, even as psychology has tended to generalize from relatively small and unrepresentative samples.[116] So, rather than claiming that science is value-free, we need to find realistic ways of acknowledging and improving this reality.[117]

A starting point is for behavioral scientists to cultivate self-scrutiny by querying how their identities and experiences contribute to their stance on a topic. Hypothesis generation could particularly benefit from this exercise, since arguably it is closely informed by the researcher's personal priorities and preferences.[118] Behavioral scientists could be actively reflecting on interventions in progress, including what factors are contributing to power dynamics.[119]

Self-scrutiny may not be enough. We should also find ways for people to judge researchers and decide whether they want to participate in research - going beyond consent forms. Finally, we should take actions to increase diversity (of several kinds) among behavioral scientists, teams, collaborations, and institutions. These could include increased support for starting and completing PhDs, reducing the significant racial gaps present in much public funding of research, or building professional networks that connect the Global North and Global South.[120]

Full detail on page 91

CONCLUSION

When considered together, these proposals present a consistent and coherent vision for the future of applied behavioral science. A common theme throughout the ten proposals is the need for self-reflective practice. In other words, a main priority for behavioral scientists is to recognize the various ways that their own behavior is being shaped by structural, institutional, environmental, and cognitive factors.

However, as the field itself shows, a gap often emerges between intention and action. Given what's at stake, BIT will focus on bridging this gap in the coming years. Realizing these proposals will require sustained work and experiencing the discomfort of disrupting what may have become familiar and comfortable practices. Indeed, this manifesto forms part of a new collection of resources from BIT to start to fulfill the goals set out here.

Improving applied behavioral science has some characteristics of a social dilemma - benefits are diffused across the field as a whole, while costs fall on any individual party who chooses to act (or act first). Practitioners are often in competition. Academics often want to establish a distinctive research agenda. Commissioners are often rewarded for risk aversion. Impaired coordination is particularly problematic, since it forms the basis for several necessary actions (such as the multi-site studies to measure heterogeneity).

Solving these problems will be hard. Funders need to find mechanisms that adequately reward coordination and collaboration by recognizing the true costs involved. Practitioners need to perceive the competitive advantage from adopting new practices and be able to communicate them to clients. Stepping back, the starting point for these changes needs to be a change in the narrative about what the field does and could do. The "manifesto" presented here aims to help shape this narrative.

Category	Proposal	Recommended action(s)
Scope	**Use behavioral science as a lens**	Present behavioral science as a lens that improves the view of any public and private issue, in order to break a self-sustaining pattern that has directed behavioral science away from the most significant problems.
	Build behavioral science into organizations	Focus less on how to set up a dedicated behavioral science team, and more on how the approach can be integrated into an organization's standard processes by upgrading its "choice infrastructure".
	See the system	Use aspects of complexity thinking to improve behavioral science so it can: exploit "leverage points"; model the collective implications of heuristics; alter specific features of systems to create wider changes; and understand the longer-term impact on a system of a collection of policies with varying goals.

Category	Proposal	Recommended action(s)
Methods	Put RCTs in their place	Strengthen RCTs to deal better with complexity by: gaining a better understanding of the system interactions and anticipate how they may play out; setting up RCTs to measure diffusion and contagion in networks; building feedback and adaptation into the design of RCTs and interventions.
	Replication, variation, adaptation	Identify the most reliable interventions, develop an accurate sense of the likely size of their effects, and avoid the weaker options. Recognize that heterogeneity requires a much higher bar for claiming that an effect holds true across many unspecified settings. Create multi-site studies to systematically study heterogeneity in a wider range of contexts and participants. Codify and cultivate the practical skills that successfully adapt interventions to new contexts.
	Beyond lists of biases	Emphasize theories that are "practical": they fill the gap between high-level frameworks and jumbled lists of biases; they are based on data and generate testable hypotheses, but also specify the conditions under which a prediction applies; they present actionable steps to solve real-world problems.
	Predict and adjust	Develop the practice of getting behavioral scientists to predict the results of experiments, and then feeding back the results to them.
Values	Be humble, explore, and enable	Avoid using the term "irrationality"; practice "epistemic humility"; and design processes and institutions to counteract overconfidence. Pay greater attention to people's own interpretations of their beliefs, feelings, and behaviors. Reach a wider range of experiences, including marginalized voices and communities. Recognize how apparently universal cognitive processes are shaped by specific contexts. Use six criteria to assess when to enable people to use behavioral science themselves.
	Data science for equity	Use data science to identify the ways in which an intervention or situation appears to increase inequalities and introduce features to reduce them. For example, groups that are particularly likely to miss a filing requirement could be offered pre-emptive help.
	No "view from nowhere"	Cultivate self-scrutiny; find new ways for the subjects of research to judge researchers; take actions to increase diversity among behavioral scientists and their teams, such as building professional networks between the Global North and Global South.

SCOPE

01

USE BEHAVIORAL SCIENCE AS A LENS

> We need to see behavioral science as a lens that improves the view of any public and private issue, rather than as a tool that we sometimes pick up. Making this change will help break the self-sustaining pattern whereby demand for behavioral science, and the tools we have developed, has pushed work towards 'downstream' interventions and away from structural changes.

The recent surge in applying behavioral science to practical issues has made a measurable difference across many domains. The approach has been adopted by public sector bodies at the local, national, and supra-national level,[121] and by private companies large and small. They have improved outcomes in health,[122] education,[123] sustainability,[124] transport,[125] diet,[126] and financial behavior,[127] among many areas. Many of these improvements have come at relatively low cost.[128]

Despite these achievements, objections have emerged. A common one is that there's been a focus on tractable and easy-to-measure changes, at the expense of bigger impact on major issues. Behavioral science, it's claimed, has just been tinkering around the edges of fundamental problems.[129]

We agree with the challenge that behavioral science can and should do more. Every day, new policies cut against well-established evidence of how people behave.[130] Services are shaped in ways that people cannot navigate. Products are launched with fundamental misconceptions about how people are likely to approach them. There are fewer prominent examples that clearly show how governing policies and systems have been designed using concepts from behavioral science, as opposed to specific aspects of how those policies were presented or structured.[131]

So, how to move forward? Step one is to realize that the strengths that have brought success may also be holding behavioral science back.
To explain, let's go back to the start of the current phase of applied behavioral science (around 2008-2012), when there was a pressing need to demonstrate clear results and build credibility. That pressure led us and others to form standard ways of applying "behavioral insights". These approaches generally have a common set of features.[132] The standard series of actions looks something like this:

- scoping the issue and exploring drivers of behavior
- defining a specific target behavior that can be measured reliably
- generating evidence-based interventions to change the target behavior
- creating a robust experimental design to test the intervention's effects
- if desired, taking the intervention to new places (e.g., "scaling")

These actions are usually presented in a step-by-step (linear) way, although most guides stress that people can loop between stages.

Over the past decade, this kind of approach has tended to produce:

- "downstream" interventions that concern how specific aspects of a policy, product or service are designed
- a focus on discrete behaviors by actors (e.g., people, businesses), considered mostly in isolation.[133]

Perhaps a good example is BIT's project to reduce missed hospital appointments in the UK.[134] This work identified the wording of text message reminders as an opportunity for improvement. We ran two randomized controlled trials in London, which found that a message referring to the cost of a missed appointment for the health system reduced no-shows from 11.1% to 8.5% - a 25% relative change. This low-cost change was then taken up by other health providers around the world.[135]

As this example shows, the approach is a strong one. There's a neatly-defined problem, a specific intervention, and a strong causal link between that intervention and the target outcome. There's a clear story to tell about what happened and why. These strengths mean that there are still so many improvements that this approach could achieve. For instance, one priority should be to clear the vast administrative burdens (or 'sludge') that prevent people - particularly those with fewest resources - from understanding or accessing government services.[136]

However, the justified focus on these clear and credible results about downstream impact also becomes self-reinforcing. People start to think that this is the sole way that behavioral science can be applied. In turn, this perception shapes demand: only certain kinds of problems are seen as ones 'suitable' for behavioral scientists.[137]

Opportunities, skills, and ambitions have been constricted as a result. In general, practitioners have focused more on expanding the application of some "tried and tested" interventions to new areas, and less on exploring new ones (or getting a deeper understanding of familiar ones).[138]

We need a rebalancing. Behavioral science also has much to say about broader, larger issues in society like discrimination, pollution, or economic mobility, and the structures that produce them. Behavioral science has the potential to fundamentally change how we understand the factors shaping behavior and therefore how we constitute an issue and what is possible.[139]

Take the economy: behavioral science can show how to regulate markets differently;[140] how to design taxes to drive wider behavioral changes;[141] and even offer a vision for the future "behavioral economy" as a whole.[142]

As this list shows, there are examples of how behavioral science has tackled more complex, structural, "upstream" issues. But these examples are harder to communicate because they often deal with the fluid, murky, and fractured narratives of politics and policy-making. Unlike the clear, linear stories of the approach outlined above, the contribution of behavioral science may be difficult to trace or may play out over a long timeframe. It's easier to talk about the neat narratives instead, particularly since many people are aware of and curious about the idea of 'nudging', which is often associated (inaccurately) with small presentational tweaks only.[143]

The wide potential scope of applied behavioral science is an idea that BIT has promoted consistently since its creation.[144] But the self-reinforcing limiting factors we outlined have proved strong. Now the increasingly urgent question is: how can we successfully change behavioral science itself?

Our answer is to consider the proposals in this manifesto, which aim to create a package that can achieve that change. We can start by trying to switch the metaphors or frames through which we perceive behavioral science itself.

Behavioral science should be understood as a lens that enhances the view of any public and private issue, rather than as a tool that we sometimes pick up.

The trends we highlight above have tended to reinforce this tool metaphor, which encourages this way of thinking:

Behavioral science is a specialist tool that is applied to certain kinds of problems - and not others. Often these are defined "delivery" issues ("How do we structure this message?"), but sometimes it can be used to solve a "behavioral" problem as an alternative to more traditional approaches like rules and incentives.

These implications lead us down the wrong path. Instead, behavioral science should be understood as a lens that can be applied to any public and private action. This change offers several advantages:

- A lens metaphor shows that behavioral insights can enhance the way we see policy options (for example, revealing new ways of structuring taxes), rather than just acting as an alternative to them.
- A lens metaphor conveys that the uses of behavioral insights are not limited to creating new interventions. A behavioral science lens can, for example, help reassess existing actions and understand how they may have unintended effects. It emphasizes the behavioral diagnosis of a situation or issue, rather than pushing too soon to define a precise target outcome and intervention.[145]

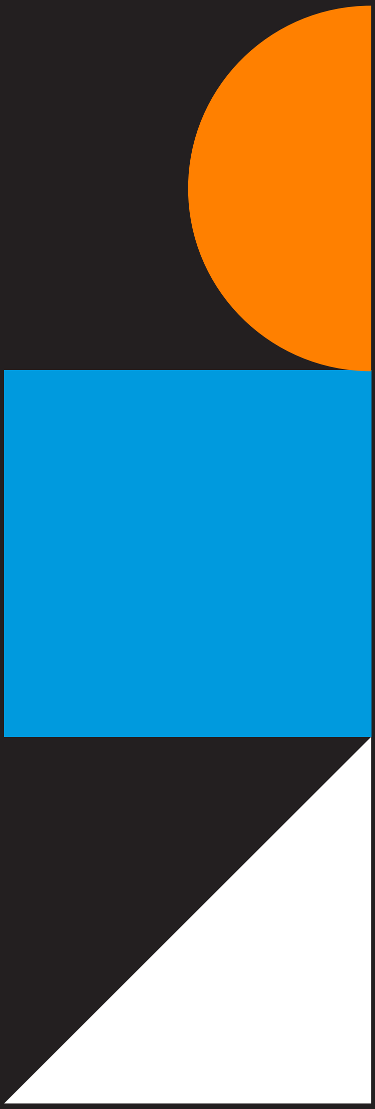
- Specifying that this lens can be applied to any action conveys the error of separating out "behavioral" and "non-behavioral" issues: most of the goals of private and public action depend on certain behaviors happening (or not). Behavioral science should therefore be integrated into an organization's core activities, rather than acting as an optional specialist tool.[146]

In one sense, this proposal is about returning to first principles. Back in 2010 we emphasized that 'civil servants...need to better understand the behavioral dimension of their policies and actions', and also stressed how behavioral science 'powerfully complements and improves conventional policy tools'.[147] But, for all the reasons above, this aspect has been less prominent over the last decade.[148]

Other metaphors apart from a lens would be powerful as well. For example, moving from 'choice architecture' to 'choice infrastructure' effectively highlights the broader, embedded nature of our behaviors.[149] The point is that behavioral science itself shows us the power of framing: the metaphors we use shape the way we behave, and therefore can be agents of change.[150]

Metaphors are particularly important because the task of broadening the use of behavioral science requires making a compelling case to decision makers. Behavioral science practitioners need to understand their audience and then shape their offers accordingly.

In a way, the metaphor of behavioral science as a tool that produces clear results has served the field well over the last decade - it has established credibility and acceptance in a defined area. The challenge now is to expand beyond that area, allowing behavioral science to fulfill its potential before the self-reinforcing cycle becomes too hard to break.



02 BUILD BEHAVIORAL SCIENCE INTO ORGANIZATIONS

> There has been too little focus on using behavioral science to shape organizations themselves, as opposed to increasing how much an organization uses behavioral science to achieve its goals. We need to talk less on how to set up a dedicated behavioral function, and more about how behavioral science can be integrated into an organization's standard processes.

For example, as well as trying to ensure that a departmental budget includes provisions for behavioral science, why not use behavioral science to improve the way this budget is created (e.g., are managers anchored to outdated spending assumptions)?

But we need to understand how this new way of thinking maps against the existing debate about how to set up a behavioral function in organizations. We propose that doing so reveals six main scenarios.

- In the **"Baseline"** scenario there is limited awareness of behavioral science in the organization, and its principles are not incorporated into processes.
- In the **"Nudged Organization,"** levels of behavioral science awareness are still low, but its principles have been used to redesign processes to create better outcomes for staff or service users.
- In **"Proactive Consultancy,"** leaders may have set up a dedicated behavioral team without grafting it onto the organization's standard processes. This lack of institutional grounding puts the team in a less resilient position, meaning it must always search for new work.
- In **"Call For Experts,"** an organization has concentrated behavioral expertise, but there are also prompts and resources that allow this expertise to be integrated more into "business as usual". Expertise is not widespread, but access to it is. This setup could mean that processes stimulate demand for behavioral expertise that the central team can fulfill.
- In **"Behavioral Entrepreneurs,"** there is behavioral science capacity distributed throughout the organization, either through direct capacity building or recruitment. The problem is that organizational processes do not support these individual pockets of knowledge.
- Finally, a **"Behaviorally-Enabled Organization"** is one where there is knowledge of behavioral science diffused throughout the organization, which also has processes that reflect this knowledge and support its deployment.

The common success factor in these scenarios is an upgrade of the "choice infrastructure" of organizations. To do this, we propose: reducing the costs of experimentation, creating a system that can learn from its actions; and developing new and better ways of using behavioral science knowledge to analyze the behavioral effects of processes, rules, incentives, metrics, and guidelines.

The second proposal is to broaden the scope of how behavioral science is used in organizations. Much attention has been paid to how the practice of applying behavioral science can have more influence within organizations – usually by advising on how a dedicated behavioral science function should be structured.[151]

This is an important question: there were more than 300 such units by 2020 worldwide; BIT and others have advised on setting them up.[152] Work in this area has covered important questions like how to arrange the leadership, organizational structure, funding, and goals of behavioral insights teams for maximum success.[153]

In contrast, there has been less attention paid to how behavioral science can be integrated into an organization's own processes.[154] There has not been enough focus on using behavioral science to shape organizations themselves, as opposed to increasing how much an organization uses behavioral science to achieve its goals.

For example, as well as trying to ensure that a departmental budget includes provisions for behavioral science, why not use behavioral science to improve the way this budget is created (e.g., are managers anchored to outdated spending assumptions)?[155]

The overriding message here is for greater focus on the organizational changes that indirectly apply or support behavioral science principles, rather than just thinking through how the direct and overt use of behavioral science can be promoted in an organization. There are two main advantages to doing this:

SCALE

Building behavioral science into organizations can address some of the issues surrounding successful scaling of interventions.[156] If some of the barriers to scaling concern cognitive biases in organizations, these changes could minimize the effect of such biases.[157] Rather than starting with a behavioral science project and then trying to scale it, we could start by looking at operations at scale and understand how they can be influenced. There is a change of perspective here to a position where 'what is scalable is not the content of what is learned in any given context, but the capacity for learning itself'.[158]

Greater potential for scale

		BEHAVIORAL SCIENCE KNOWLEDGE AND CAPACITY		
		LIMITED	CONCENTRATED	DIFFUSED
BEHAVIORAL SCIENCE INCORPORATED INTO ORGANIZATIONAL PROCESSES	NO	Baseline	Proactive consultancy	Behavioral entrepreneurs
	YES	Nudged organization	“Call for the experts”	Behaviorally-enabled organization

RESILIENCE

The goal would be to produce behaviorally-informed standard or “business as usual” processes, rather than the continued application of behavioral science explicitly. That approach is resilient to changes in the demand for “behavioral science” solutions as such in the future.

BIT’s 2018 Behavioral Government report proposes many practical changes to organizational processes to mitigate biases in government.[159] But we need to understand the range of options for implementing such changes. How do we think about them alongside the desire to create a dedicated behavioral insights team, for example?

We think that the diagram above offers a useful way of mapping the options for building behavioral science into organizations.[160]

The vertical axis represents whether behavioral science has been used to shape the organization’s own structures or processes, using a crude yes/no distinction to make the diagram manageable. We will bring this distinction to life with examples in the following sections, before defining it in more detail.

The horizontal axis deals with the extent and form of behavioral science knowledge and capacity in an organization.[161] In the ‘Baseline’ scenario, there is very little or no awareness of behavioral science concepts in the organization. ‘Concentrated’ refers to the setup where there is a dedicated team or resource that applies behavioral science to organizational priorities.[162] In the ‘Diffused’ scenario, people or teams with competence in behavioral science are spread throughout the organization. Deciding between concentrated and diffused setups is generally seen as a central choice for organizations looking to build a behavioral science function.[163]

We can now see how this framework illuminates the different options:

BASELINE

Here, there is limited awareness of behavioral science in the organization, and its principles are not incorporated into processes. Benefits are likely to be limited.

NUDGED ORGANIZATION

Here, levels of behavioral science awareness are still low, but its principles have been used to redesign processes to create better outcomes for staff or service users. For example, the pervasive optimism bias in organizations' plans can be reduced by mandatory 'pre-mortems', where decision makers imagine the future failure of their project and then work back to identify why things went wrong. Group reinforcement (or "groupthink") could be minimized by creating routes for diverse views to be fed in anonymously before and after group discussions, counteracting the pressure to conform face to face.[164] And there are vast opportunities to reduce the administrative burdens, or 'sludge', that make services and processes difficult to access and navigate.[165]

This scenario is termed the 'nudged organization' because no explicit behavioral science knowledge or capacity is created or needed. Like for nudging, it is the choice architecture (or choice infrastructure) that produces the outcomes, and there is no neutral choice in the way that an organization's processes are set up. That means no behavioral team or unit is created; the change or goals may not even be framed in terms of behavioral science (as for "administrative burdens")

For this reason, the best starting point is to understand how the existing setup is influencing behavior. Where is the choice architecture currently working well, through accident or design? How can existing processes be amended easily to draw on these practices? Who are the people who oversee the rules, incentives, metrics, and guidelines that influence people throughout the organization?

To give a concrete example, human resource leaders profoundly shape what organizations permit and reward. Yet, there has been relatively little focus on "behavioral HR". Recent studies have shown that cognitive biases such as decoy effects, framing effects, anchoring and halo effects can be created in practical decisions such as procurement and performance appraisal.[166] They can also be countered: when considering the purchase of an email software, framing effects like saying 20% of users were dissatisfied significantly affected intentions (versus saying 80% were satisfied), but these effects were eliminated if both percentages were shown (in a random order).[167]

The big outstanding question in this scenario is who introduces the nudges, since the organization has little internal capacity. Perhaps these could be one-off changes introduced from outside? Answering this question feels important, since the return on investment here could be large - and, for that reason, this model feels like a neglected opportunity that needs more attention.

PROACTIVE CONSULTANCY

In this situation, leaders may have set up a dedicated behavioral team, but perhaps not given much thought to supportive organizational changes. The result is that the team has to work in an enterprising way, going to look for opportunities and having to prove its worth.[168]

This situation reflects the reality for many teams, who are 'looking to develop networks, positions, and tactics that establish their authority and credibility among decision makers.'[169] As a result, much of the discussion has focused on how best to set up these teams. The better contributions have recognized that this question is fundamentally political, rather than technocratic - how do the people leading such a resource build relationships and present their team as useful to their organizations?

The problem with this scenario is that teams may not be in a resilient position, since they lack ways to be grafted onto the standard processes of an organization. For example, leaders may neglect to support and resource evidence-gathering and experimentation. At the same time, they may have unrealistic expectations because they know only the highlights of previous behavioral science success stories.[170]

Teams will therefore have to continually prove their worth and relevance, while having fewer options for doing so. As current practitioners point out, 'interventions that seem relatively easy to implement (e.g., an RCT with promotion letters) can require a set of system changes that touch a variety of groups in the organization (e.g., printing services, database administrators, processing centers).'[171] This kind of broader organizational scaffolding is not always prioritized, despite being needed for behavioral teams to fulfill their potential.

CALL FOR EXPERTS

In the Call for Experts scenario, an organization has similarly concentrated behavioral expertise, but there are also prompts and resources that allow this expertise to be integrated more into 'business as usual'. At its simplest, this might mean that standard procedures prompt staff to recognize that the expertise may be needed (e.g., any new requirement in an application process needs to be assessed for its likely effects on behavior). Expertise is not widespread, but access to it is.

At another level, the organization may have invested to ensure that the ability to randomize has been built into new and existing delivery systems, thereby allowing the team to run experiments when they are called on. If working well, this setup would mean that processes stimulate demand for behavioral expertise that the central team can fulfill. That team may also have the institutional support to proactively monitor activities and respond quickly to specific crises.

One benefit to this kind of setup is that it allows teams to select the most promising collaborations, rather than taking whatever is on offer. For example, the team in Employment and Social Development Canada's Innovation Lab claims that a 'careful selection process is critical to the success of incorporating behavioral insights into an organization', since it identifies partners who are open and willing to innovate, which makes it more likely that the subsequent project demonstrates the true value behavioral science can add.[172]

BEHAVIORAL ENTREPRENEURS

In this situation, there is behavioral science capacity distributed throughout the organization, either through direct capacity building or recruitment. The distribution of expertise can work if there are effective support networks and efforts at coordination.[173]

The problem with the behavioral entrepreneurs scenario is that organizational processes do not support these individual pockets of knowledge. Therefore those with expertise find it hard to apply ideas in practice, evaluate their effects, share findings, and build learning. For example, reducing "sludge" often 'requires coordination among a number of teams in the organization', which is a problem when teams work in silos and 'it is not acceptable in the organization's culture to interfere with other teams' affairs.'[174]

These are not just hypotheticals. A review of the Dutch policy landscape found that 'most behavioral policy practices have not been deeply institutionalized', and their advancement 'depends on the ambition of individual enthusiasts' (or, as we've called them, "behavioral entrepreneurs").[175] While they can achieve some successes, their lack of institutional grounding can mean that they become jaded and start looking for other options instead.

BEHAVIORALLY-ENABLED ORGANIZATION

We see a behaviorally-enabled organization as one where there is knowledge of behavioral science diffused throughout the organization, which also has processes that reflect this knowledge and support its deployment.[176] This is the most resilient setup, since staff will be applying behavioral science in a deliberate way as part of "business as usual", rather than through special projects.

A behaviorally-enabled organization would bring together some of our previous proposals. It would embed the behavioral lens mentioned earlier into its core functions, including strategy and operations. For example, behavioral science could be integrated into cross-cutting frameworks like the WHO's 'Health In All Policies' to prompt inter-departmental working.[177] It would address the need to "see the system", recognizing that sustainable outcomes are difficult to achieve through isolated changes. Such an organization would be self-reflective, and carefully explore the varying perspectives and experiences of its staff and service users.

While this setup has the greatest opportunity for scale and sustainability, it also requires the greatest investment. We conclude by talking about what kinds of investments are needed.

CHOICES AND PRIORITIES FOR THE BUILD

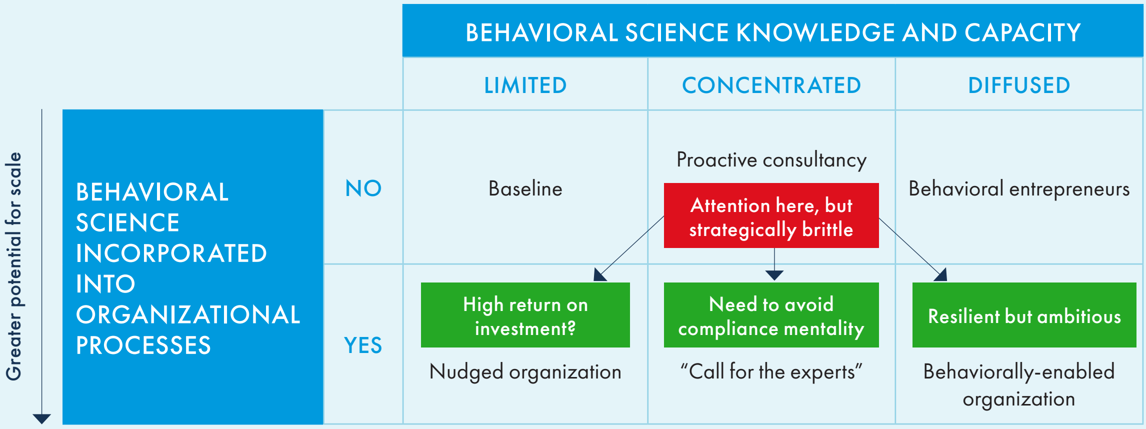
Most discussions make it seem like the meaningful choice is between the different columns in our framework - how to organize your dedicated behavioral science resources. We argue that the more important move is from the top row to the bottom row: moving from projects to processes, from commissions to culture.

A useful way of thinking about this task is about building or upgrading the "choice infrastructure" of the organization, defined as 'the institutional conditions and mechanics of systems - the structures, processes, and capabilities - that directly underlay and support behavioral interventions to help choice architecture solutions work effectively and as planned'.[178]

In other words, we should place greater focus on the institutional conditions and connections that support the direct and indirect ways that behavioral science can infuse organizations. As the diagram below highlights, there are choices to be made about how this is done, based on ambitions and resources.[179]

Working out how best to build the choice infrastructure in organizations should be a major priority for behavioral science. As with many systems, the best option may be to focus on creating the conditions for desired behaviors to emerge, rather than over specifying solutions.[180] But already we can see some features will be crucial.
Dilip Soman and Katherine Yeung argue for the importance of reducing the costs of experimentation, including cheaper data collection, creating an experimental mindset, reducing institutional impatience, and building agility so that organizations can easily adapt to learning.[181] Others have promoted the importance of sharing learning itself, pointing towards the crucial role of Singapore's Civil Service College in 'curating and facilitating an ecosystem of learning opportunities' for behavioral science.[182]

To this list, we want to add new and better ways of using behavioral science knowledge to analyze the behavioral effects of processes, rules, incentives, metrics, and guidelines. Such work has surged recently under the labels of 'behavioral public administration' and 'behavioral operations management', building on a longer tradition of organizational behavior research.[183] We need to ensure that this agenda produces work that has practical value (and not just for the public sector), as in the proposal of "sludge audits" to reduce administrative burdens.[184] Doing so will mean that the behavioral lens we just proposed can be used by an organization's members - and, ideally, by its leaders, who are in a position to achieve broader, systemic change.



		BEHAVIORAL SCIENCE KNOWLEDGE AND CAPACITY		
		LIMITED	CONCENTRATED	DIFFUSED
BEHAVIORAL SCIENCE INCORPORATED INTO ORGANIZATIONAL PROCESSES	NO	Baseline	Proactive consultancy Attention here, but strategically brittle	Behavioral entrepreneurs
	YES	High return on investment? Nudged organization	Need to avoid compliance mentality "Call for the experts"	Resilient but ambitious Behaviorally-enabled organization

03 SEE THE SYSTEM

> Many big policy challenges emerge from complex adaptive systems, which present major challenges to the dominant way that behavioral science has been applied. However, we can adapt behavioral science to deal with complexity better, and use it to: identify "leverage points" where a specific shift in behavior will produce wider system effects; understand the collective implications of individuals using simple heuristics to navigate a system; and change the rules of that system to make it more likely that desired behaviors will emerge.

Of course, not every problem will involve a complex adaptive system. So behavioral scientists should first develop the skills to recognize the type of system that they are facing ("see the system"), and then choose their approach accordingly. Fulfilling the broader promise of behavioral science also requires us to expand the ways we tackle problems. The process of identifying targets, exploring drivers, developing solutions, and testing them is strong. The problem is that it contains several assumptions that do not hold when confronting some of the biggest challenges societies face. That's because these challenges often consist of behaviors in complex adaptive systems (CAS).[185]

The risk of talking about 'complexity' is that it may seem like just another way of saying problems are difficult (indeed some use the term as an excuse to do nothing).[186] In fact, we mean applying a particular way of analyzing the world. Our proposal is that combining behavioral science with complexity thinking offers new, credible, practical ways of doing things differently.

First we need to explain complex adaptive systems briefly:

"A complex adaptive system is a dynamic network of many agents who each act according to individual strategies or routines and have many connections with each other. They are constantly both acting and reacting to what others are doing, while also adapting to the environment they find themselves in. Because actors are so interrelated, changes are not linear or straightforward: Small changes can cascade into big consequences; equally, major efforts can produce little apparent change. An important point is that coherent behavior can emerge from these interactions— the system as a whole can produce something more than the sum of its parts."[187]

CAS consist of many different causes, actors, and goals. There are many examples of them in human societies, including cities, markets, criminal justice systems, and political movements. They often create what have been called 'wicked problems', which are ''difficult or impossible to solve because of incomplete, contradictory, and changing requirements that are often difficult to recognize".[188]

Such problems even produce little agreement among different groups about how 'success' can be defined.

The need to understand CAS is becoming more urgent. Complexity has emerged from the sheer number of contacts created by global population growth. At the same time, technological changes like the growth of social media mean that information can be transmitted much faster and cheaply, in an unchanged form, over many different geographies and networks. The fact we do not fully understand how these changes affect human behavior is 'a principal challenge to scientific progress, democracy, and actions to address global crises'.[189]

Indeed, the Covid-19 pandemic, perhaps the most acute global policy challenge of recent times, has several features of a wicked problem:[190]

CONTESTED IDEAS OF SUCCESS

Throughout the pandemic, there has been disagreement between individuals, organizations, and countries about what the overall policy goals should be (suppression, elimination, buying time until vaccines, removing restrictions), let alone consensus about how to achieve those varying goals.[191]

NON-LINEARITY

The initial coronavirus variant appears to have been 'over-dispersed' - outbreaks were seeded by a handful of super-spreading events.[192] It's estimated that around 80% of infections were caused by just 10% of individuals.[193] These properties can explain why some nascent outbreaks fizzle and others take off: viral spread is a non-linear process.[194]

Understanding these features could target policy responses effectively. They might indicate that stopping many repeated contacts within a small set of individuals has little effect on spread, in contrast to reducing random contacts at events and restaurants.[195] In other words, the way the virus interacts with social systems means preventing many social contacts is unlikely to reduce viral spread in a linear way, whereas preventing a few super-spreading events may have an outsized impact.

UNINTENDED CONSEQUENCES

The wide-ranging, intensive actions taken to mitigate the spread of Covid-19 have had many indirect effects. These may have included increases in domestic violence; shortages of hydroxychloroquine in West Africa; reductions in carbon emissions; and changes in healthcare access with shifts to telehealth.[196] Our point is not that these actions should not have been taken per se, but that solutions may change the nature of a problem or create many new ones.

These ideas challenge the assumptions underlying the dominant behavioral science approach

The issue is that 'there are fewer examples of behavioral insights applied to understand behavior in complex change processes'.[197]

Why? Because the realities of complex adaptive systems challenge the main assumptions underlying the dominant behavioral science approach: tight focus on a target behavior, linear effects, and stability.[198] We outline each of these before offering a way forward.

TIGHT FOCUS ON A TARGET BEHAVIOR

People can agree on a specific, measurable target behavior. Interventions that shift this behavior are successful - wider effects are minor and may not be considered, unless pre-specified.

Some behavioral science organizations focus on '"breaking problems down into their constituent parts to understand the desired behaviours'".[199] For some issues, there is value in doing this to identify a core target behavior to drive improvement. But you cannot understand a complex system by breaking it down into parts and mapping it - the way its connections function is key.[200]

We cannot assume that changing a particular part of the system will have the desired overall outcome. For a start, there may be intense disagreement between parties about how a behavior contributes to an issue: some people may see pre-school provision as key to regenerating an urban area; others may view that as an unimportant contributor, compared with reducing crime.

Moreover, focusing on a single behavior to achieve a specific outcome may disguise unintended consequences that hinder progress towards the larger goal. A tight focus on target behaviors can mean seeing only a slice of the picture, and ignoring how actors in a system may respond by producing effects felt elsewhere.[201]

There are many such examples. In the US, grocery stores participating in a government voucher program successfully reduced fraud, but this effort also led to many stores no longer stocking high nutrient foods as a result - ultimately harming recipients' diets.[202] A ban on plastic carryout bags led to 40 million fewer pounds of plastic being used for this purpose, but 12 million more pounds purchased as large trash bags instead.[203] In Brazil, reminders for upcoming credit card payments reduced late payment fees by 14%, but also increased overdraft fees in bank accounts by 9%.[204] Increasing efficiencies in hospitals produces situations where 'activities in one area of the hospital become critically dependent on seemingly insignificant events in seemingly distant areas'.[205]

LINEAR EFFECTS

The intervention affects participants in a direct and linear way, according to a pre-developed theory of change.[206]

In a CAS, actors adapt to the behavior of others, so there is often not a simple relationship between inputs and outputs. Actors in a system may adapt to 'buffer' the effect of an attempted change and keep things apparently stable - i.e., making it seem that the intervention had no effects.[207]

However, repeated efforts may weaken these stabilizing factors, and then a minor subsequent event produces a 'tipping point', where change happens suddenly and the system flips into a new state.[208] An example might be repeated challenges that weaken the commitment of a country's armed forces to democracy, which do not translate into action but which create the conditions for an apparently minor event to trigger a coup.

The point here is that if you do not see your expected behavior in a certain timeframe, in line with a linear 'theory of change', you may assume that it has failed. But, in fact, change may be happening through routes and over timescales you had not anticipated.

STABILITY

You can measure the pre-specified target behavior between point A and point B. The system will remain stable over that time, and people will not adapt in response to the intervention.

Since actors adapt to new conditions, and influence each other in doing so, the nature of the problem may be changed by the introduction of an apparent solution itself.[209] A snapshot of behaviors at one point in time is not enough to claim victory. Perhaps the best example is regulation: market players experiment with and gradually adapt to a regulatory regime, working out how to evade its provisions, until a new policy is needed.[210] Therefore success may actually lie in how well behavioral scientists adapt to the unanticipated effects their own actions produce.[211]

What's particularly sobering for behavioral scientists is that their own field provides reasons why they may be holding onto these principles, despite their limitations. One is the 'reductive tendency', which is 'a process through which individuals simplify complex systems into cognitively manageable representations... when faced with complex concepts, individuals are often inclined to treat dynamic concepts as static, or to generalize across dissimilar domains'.[212]

In other words, behavioral scientists use heuristics based on linear models of change when confronted with complexity, since doing so is 'predictable, comforting, and less mentally taxing'.[213] This tendency can exacerbate another one - an illusion of control. Those who are creating and implementing interventions may overestimate how much control they have over events and outcomes, since they are using an inaccurate mental model of how things work.[214]

The end result, some argue, is a failure to understand how actors are acting and reacting in a complex system that leads policymakers to conclude they are being 'irrational' - and then actually disrupt the system in misguided attempts to correct perceived biases or inefficiencies.[215]

DEVELOP BEHAVIORAL SCIENCE THAT CAN TACKLE PROBLEMS IN COMPLEX SYSTEMS

We have outlined the criticisms. Now, we want to offer hope and a way forward. There is an opportunity to develop behavioral science so it can tackle the aspects of complexity that are common to major policy issues. The starting point is to show how behavioral science can shed new light on well-known features of CAS: the fact that small changes can have big impacts, and the way that actors often use a simple set of rules to navigate a system.

The idea that 'small changes can have a big impact' is often used in the sense that apparently minor features of the way a choice is designed or presented can have a surprisingly large effect on subsequent behavior. Used this way, the idea fits neatly with the standard approach of specific, isolated changes being applied to change a pre-defined behavior.

But CASs show that the statement is true in a different way. They show that 'higher-level' features of a system can actually emerge from the 'lower-level' interactions of actors participating in the system.[216] When they become the governing features of the system, they then shape the 'lower-level' behavior until some other aspect emerges, and the fluctuations continue.

Let's make things tangible with examples. In the Covid-19 pandemic, people were trying to achieve their goals (live their lives) within a broad set of rules and in response to changing events. These adaptive behaviors in specific contexts interacted with the adaptive abilities of the virus. New variants emerged as a result.[217] Some of these variants quickly became widespread and, in doing so, changed the nature of the whole pandemic, most obviously through their greater transmissibility or resistance to vaccines. Policymakers attempting to handle the changed situation then had to come up with new overall strategies.

Experiments have also shown how initial, random fluctuations can emerge and solidify into stark divides that shape societies. For example, a recent US study showed how partisan policy divisions may actually be produced by these random fluctuations.[218]
The experiment created online 'worlds' where self-identified Democrats or Republicans were asked whether they agreed with up to 20 statements. These statements concerned public issues, but had been selected so they did not reflect pre-existing partisan positions - e.g., whether there should be a move to professional full-time jurors.

In eight of the worlds, participants could see if mainly Democrats or Republicans were agreeing with the proposal. When this happened, strong partisan alignment quickly followed. But the important point is that which proposals fell into the Democratic or Republican camp varied greatly between worlds - sometimes the juror proposal was adopted by one side, sometimes the other. There was initial fluctuation in the initial stages, driven by chance and context, before a sudden non-linear alignment one way or another.

More fundamentally, we can see that norms, rules, practices, and culture itself can emerge from aggregated social interactions. These features then shape cognition and behavioral patterns in turn.[219] The implications here disrupt the crude distinctions of 'upstream' versus 'downstream' or 'high-level' versus 'low-level' policies - or, as one recent paper put it, the "individual frame" and the "system frame".[220] Instead we have 'cross-scale behaviors',[221] where behaviors embedded in specific contexts, shaped by the overall way the system functions, can self-organize and emerge to shape the system itself.[222]

This way of thinking opens up new possibilities. Behavioral science could be used to identify 'leverage points' where behavior could be nudged in a way that produces wider system effects.[223]

Perhaps these interventions could be targeted at the stage of random fluctuations, when contingent features of the context can determine which behaviors get locked in. At that stage, a small, well-timed change could shift events onto a different path. For example, a recent study shows that presenting people with a random selection of opinions (a "random dynamical nudge") from others could prevent segregated echo chambers from forming in online environments.[224]

Another option is to identify when and where 'tipping points' are likely to occur in a system - and then either nudge them to occur or not, depending on the policy goal.[225] To do this would require understanding how close the system is to abrupt change, and whether there are stabilizing forces that can be easily strengthened or weakened. To take a clear example from the animal kingdom: a study identified that a society of monkeys was near a critical point, such that just a small disturbance would lead to widespread violence. But the study also found that some individuals played a crucial role by adjudicating fights, and thus preventing a tipping point from being reached - strengthening this role could therefore prevent collapse.[226]

In the human world, a recent study showed how targeting 'structurally influential individuals' can create artificial tipping points in a similar way.[227] The same principle can be applied more widely. If even a subset of consumers decide to switch to a healthier version of a food product, this can have broader effects on a population's health through the way the system realigns. It becomes marginally more profitable to stock the healthier option, which then may change the mix of products available to consumers in general.[228]

Changing the way that the medical use of cannabis is interpreted in U.S. drug law may end up affecting a whole range of substances that are currently illegal.[229]

Behavioral science could also be used to understand when a tipping point in behaviors and attitudes is incipient and may invite or require a government response. For example, it could have identified that public attitudes and behaviors related to smoking in the UK were shifting, so that legislation banning smoking in pubs would be both respected and welcomed.[230]

But there are also limitations to focusing on tipping points. One is that potential tipping points can be difficult to identify.[231] Inefficiency and multiple attempts may be inevitable. Moreover, this approach ignores the possibility of influencing the system to get to a tipping point in the first place. Fortunately, there is another route for behavioral science to address the structural features of a system.

We've seen that the connections between different actors in a system can produce complex, multi-scale outcomes. But the individual actors in this system often rely on a core set of relatively simple rules to guide their behavior - e.g., 'do what others are doing', 'take the first available option'.[232] Seen through the lens of behavioral science, these rules are understood as mental shortcuts or frugal heuristics, and have been closely studied as the means by which people navigate their environments.[233]

We can now start to see a future approach that fuses behavioral science and complex systems, in order to:

- Provide evidence, not assumptions, about how people use heuristics to guide their interactions.
- Study the collective implications of these heuristics.
- Show how these heuristics play out against the governing conditions and structures of a system.
- Change the rules of that system to make it more likely that desired outcomes will emerge.

There are already examples that point the way. While it is known that factors like "animal spirits" and narratives affect the economy, macroeconomic models play a big role in guiding fiscal policies.[234] These models are often built on the assumption of "rational agents" which, as behavioral science shows, is not always accurate.

New studies are emerging that model households as actors embedded in shifting social networks who use heuristics in response to other actors. For example, a recent one assumes that a household 'updates its savings rate by copying the savings rate of its neighbor with the highest consumption'.[235] The study shows that if the speed at which this copying happens is changed, then a tipping point occurs and the households divide into a rich group and a poor group. In other words, modifying heuristics can produce non-linear shifts with profound effects.

What are the practical implications? The standard approach would suggest trying to directly influence the way people use heuristics - for example, by creating interventions to change the way households learn about savings from each other. But a better way could be to harness the power of the system and make a targeted change in one of its features, which then drives wider effects.

Here are some examples of what we mean.

EXAMPLE

We mentioned earlier that social media has been viewed as a major factor in driving changes in collective human behavior. A core constituting feature for social media behaviors is the ease with which information can be shared. Even minor changes to this parameter can drive widespread changes - some have argued that such a change is what created the conditions leading to the Arab Spring, for example.[236] If 'changes to a few lines of code can impact global behavioral processes', then a priority should be understanding these changes and how they can be made, if appropriate.[237]

EXAMPLE

The UK's tax on sugared drinks also worked by identifying and modifying a key system parameter. In that case, the tax was designed in tiers, so that higher levels of sugar content resulted in higher taxes. This design choice altered the incentives presented to manufacturers to make it more attractive to reformulate their products. Rather than directly persuading consumers to consume less sugar, this approach instead tries to gear the system dynamics of the market towards reductions in the sugar content of available products.[238]

EXAMPLE

In London, red lights are typically the default at pedestrian crossings. People have to push a button and wait for the green signal to go ahead. In 2021, Transport for London changed 18 crossings so that their crossing lights default to green. Pedestrians didn't have to stop and press a button. They'd see the green light and be able to cross right away - unless a vehicle was detected, at which point, the signal would turn red automatically. The change meant that pedestrians saved time, complied with signals more, and there was virtually no increase in delays for traffic.[239] This was both a classic nudge (changing the default) and an example of changing one part of a system to have wider effects. Increasing compliance with the signals makes walking safer; the default change makes walking faster. The result could be an increase in pedestrian traffic that reshapes overall transit policies and services.

In these examples, the idea is to find targeted changes to features of a system that create the conditions for wide-ranging shifts in behavior to occur. Complexity is used to structure effective action, rather than an excuse for doing nothing.

This approach also suggests that a broader change in perspective is needed. We need to realize the flaws in launching interventions in isolation and then moving on when a narrowly defined goal has been achieved. Instead, we need to see the longer-term impact on a system of a collection of different policies with varying goals.[240] Doing so will also address some of the criticism directed at behavioral economics regarding "ergodicity" - that it neglects how repeated decisions are taken over time in changing contexts, and wrongly diagnoses irrationality as a result.[241]

One way of thinking about this ambition is that behavioral scientists should be thinking more about system stewardship. A decade ago, system stewardship was proposed as a way of harnessing the power of complex adaptive systems, while recognizing that they cannot be controlled in a direct or linear way.[242]

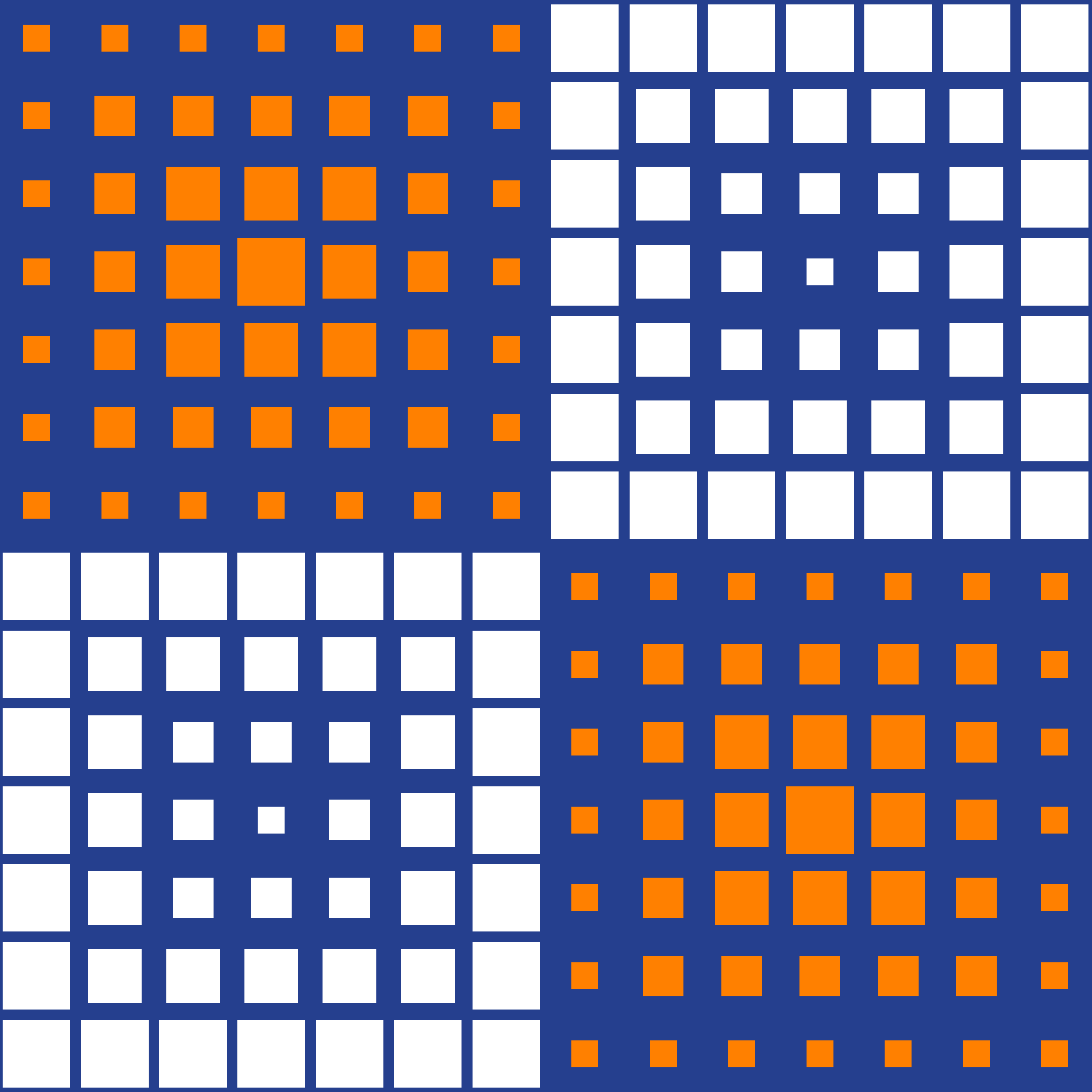
We don't explain system stewardship in detail here, but it involves setting high-level goals or a set of simple rules for actors in the system, seeking feedback, and responding to how the system is adapting as a result. The idea is to create the conditions for certain behaviors, see how a system can respond to issues through emergence and adaptation, and steer that adaptation towards broad goals if needed.

SEE THE SYSTEM

Finally, we want to turn to the main message of this proposal. Not every problem that behavioral scientists face will involve a complex adaptive system. So what practitioners should be doing is first understand the type of system that they are dealing with, and then choose their approach accordingly. You need to see the system.

When we are dealing with a simple problem or system, the linear process works well. But in situations of complexity it may produce false precision that misses what is actually going on. In those cases, you should be looking for leverage points, whether tipping points or targeted changes to system features, and taking a system stewardship perspective.

Of course, identifying the nature of the system you are dealing with may not be easy, since different elements of complexity may emerge over time.[243] But evidence is emerging that people can be trained to recognize the features of a CAS. One promising route is to share core concepts with people and then have them play agent-based simulations as, say, farmers or policy makers; these simulations can show vividly how well-intentioned changes in one area can have unintended consequences in another.[244] Others are exploring the value of novel immersive technologies,[245] or just basic checklists.[246] We think behavioral science can play a proactive role in this effort.



METHODS

04

PUT RCTs IN THEIR PLACE

> RCTs have been a core part of applied behavioral science, and they work very well in relatively simple and stable contexts. But they can fare worse in complex adaptive systems, whose many shifting connections can make it difficult to keep a control group isolated, and where a narrow focus on predetermined outcomes may neglect others that are important but difficult to predict.

We can strengthen RCTs to deal better with complexity: we can try to better anticipate indirect outcomes that may occur; to set up RCTs to measure diffusion and contagion in networks; and to adopt adaptive trial designs. We can also use behavioral science to improve alternative ways of measuring impact - in particular agent-based modeling, which often relies on assumptions of rational choice.

Where does all this leave randomized controlled trials (RCTs)? RCTs have been a central part of the applied behavioral science process, given its emphasis on quantitative methods that produce a robust counterfactual to estimate an intervention's impact.[247] Yet they may deal poorly with the features of complex adaptive systems. We can see this in two main ways: dealing with change and establishing causality.[248]

DEALING WITH CHANGE

The generally accepted best practice for RCTs is to choose specific, measurable outcomes in advance and pre-register them; changing them subsequently requires effort and may even attract suspicion.[249] But, as we pointed out, a narrow focus on predetermined outcomes risks neglecting ones that are important but difficult to predict.[250] And new outcomes are not the only issue: "new questions, causal pathways, stakeholders or even objectives may emerge during the evaluation which the original evaluation design does not accommodate".[251] The intervention itself may destabilize the system and reconfigure the nature of the issue at stake. Generally, RCTs are not good at dealing with this kind of emergent change.

ESTABLISHING CAUSALITY

One of the main advantages of RCT is that they identify causal effects by separating out a control group that is not exposed to an intervention. However, the many shifting connections in a CAS make it more difficult to ensure that a control group remains isolated. This kind of 'contamination' is a particular problem for policy makers who are tasked with influencing a system as a whole - for example, it is difficult to conduct an RCT on the introduction of new tax initiatives or criminal justice guidelines at the national level.

The way that non-linear change happens over time also presents a challenge to RCTs. For example, we may plan to measure effects over a predefined period after an intervention, say 12 weeks, based on a linear theory of change. The problem is that the change may not be linear - nothing may happen initially, but then change occurs suddenly and in a non-linear way after 3 weeks (or 13) and continues to grow after the 12 week cut off.[252] Of course, this is more a point about how many RCTs are applied in practice, rather than about what RCTs can do if they are designed perfectly.

We want to stress that, in contexts that are simpler and more stable, RCTs work well. There were very good reasons why applied behavioral science embraced the precision, skepticism, and emphasis on causality that real-world RCTs can bring. However, we also need to respond to these challenges by

- Finding ways of strengthening RCTs so they are better aligned to the demands of evaluating changes in complex systems.
- Identifying how behavioral science can use and enhance other ways of evaluating impact.

STRENGTHENING RCTS

The first opportunity to strengthen RCTs comes in the planning phase. We can try to better anticipate indirect outcomes that may occur and draw boundaries or select measures accordingly. For example, BIT worked on an RCT that tested the impact of sending parents texts encouraging them to ask their children questions about their science curriculum.[253] As intended, the intervention increased at-home conversations about science. But the RCT design also made it possible to see that the texts also made parents less likely to engage in other school-related discussions and supportive behaviors (e.g., turning off the television). By anticipating this possibility, the RCT could measure the broader shifts in behavior created by the intervention.

The need, therefore, is to gain a better understanding of the system interactions and anticipate how they may play out.[254] This can be done through "dark logic" exercises that try to trace potential harms, rather than benefits, thereby challenging assumptions about an intervention's effects.[255] One such study indicated that providing clinicians with concerning feedback on their performance may impede (rather than improve) that confidence.[256] Engaging the people who will implement and participate in an intervention will be a key part of this effort. Another potential route is to use scenarios or early-stage prototypes to gain insight into the way issues may emerge.[257]

Similar thinking has been used to develop "hybrid implementation-effectiveness designs".[258] As well as measuring overall effectiveness, these designs try to identify how the different parts of an intervention interact: for example, in the case of the education RCT mentioned, the parents, the students, and school all played different roles. These designs typically require greater stakeholder involvement in order to get good insights into these interactions, but the interventions may be adopted more widely as a result.

The second opportunity is to set up RCTs to measure diffusion and contagion in networks. An example is the study on how political partisanship forms, mentioned earlier, which created separate online "worlds" to compare how contagion and adaptation play out in different conditions. Indeed, online may be the natural home for these kinds of RCTs, since they require large populations, complete adoption data, complete network data, and replication.[259]

But they can also be done in the real world, through cluster randomizations at the network level (if those networks are not well-connected with each other). One way this has been done is by testing how social networks in different villages adapt to different interventions. For example:

EXAMPLE

Randomizing villages in Honduras to target a) randomly selected villagers b) villagers with the most social ties or c) nominated friends of random villagers, as the people to promote use of water purification and vitamin intake. The behaviors spread most widely when route C was taken.[260]

EXAMPLE

Adoption of new, higher productivity farming technology was greater in Malawian villages where 'seed farmers' had been selected by network theory-based approaches, rather than using a government extension worker to identify them.[261]

EXAMPLE

In 522 Indian villages, getting people to nominate 'gossips' led to much higher uptake of immunizations than in villages where the same information was provided to randomly selected individuals.[262]

BIT has also produced guidance on how to evaluate complex interventions affecting whole schools.[263]

The third opportunity is to build feedback and adaptation into the RCT design.[264] An obvious starting point is to actively seek out emerging change and 'weak signals' that show how the intervention is playing out in practice.[265] To use the previous example of encouraging parental conversations about science - if the researchers had not anticipated that other kinds of parental support might decline, they might have discovered this was happening by talking to students about how things were going. To handle this possibility, BIT has proposed using a 'two-stage trial protocol', where new research questions are added while a trial is in progress, but before analysis starts. This two-stage process allows emerging possibilities to be analyzed, without entering the problematic territory of "hypothesizing after results are known".[266] We can also use feedback to adapt the trial design and interventions, as in a Multiphase Optimization Strategy.[267] For example, one study aimed to improve the reading skills of third-grade students in Colombia through remedial tutorials in small groups.[268] Since three consecutive cohorts of students were involved, the results of each wave could be used to fine-tune the intervention for the next wave. As a result, the effectiveness of the intervention increased over time.

An 'evolutionary RCT' has been offered as a specific form of adaptive trial.[269] Here, participants are randomized to the treatment group at a 2:1 ratio compared to the control group. The idea is that behavioral scientists would explore how the intervention is working, run side experiments, adapt the intervention in response, and then 'split off' some of the treatment group to assign them to the adapted intervention.

A similar option is a SMART (Sequential Multiple Assignment Randomized Trial) or a micro-randomized trial.[270] The idea here is that you actively explore and monitor how people are responding to an intervention - if the evidence shows it seems to be ineffective for some, then they are re-randomized to another intervention or a control group. For example, if an initial diabetes prevention intervention does not seem to be affecting the blood glucose levels of some people, they are split off and re-randomized to a more intensive option. The idea is that you can evaluate an adaptive series of interventions.

This is a fast-developing area.[271] For example, there is much interest in the contributions machine learning can make. "Bandit" algorithms can be used to identify which interventions are working best, and gradually add more people to those experimental arms (in other words, automating the "evolutionary RCT" process).[272] These have been commonplace in the A/B marketing tests, but have some drawbacks.[273] We are still learning which conditions favor adaptive trials - e.g., when outcomes vary widely between groups; when results can be realized and observed in waves, as in school years or terms; and when it is easy to switch people between interventions.

However, adaptive designs do not address all the issues mentioned earlier. And they have disadvantages: for evolutionary RCTs, it can be hard to make the case for recruiting a bigger than normal treatment group; SMARTs can be complex to manage and are quite focused on individuals, rather than gaining system-level feedback.[274] But they represent improvements worth considering when dealing with complex adaptive systems.

USING AND ENHANCING OTHER WAYS OF EVALUATING IMPACT

There are other ways of assessing effects in complex environments as well. One approach that is particularly relevant to behavioral science is agent-based modeling (ABM). ABM tries to simulate the interactions between the different actors in a system, rather than imposing the functioning of a system in a top-down way.[275] That means ABM can replicate many of the features seen in complex systems, like tipping points and the emergence of higher-level features from lower-level features.

The practical point is that ABM can create a virtual counterfactual - a representation of what would have happened in the absence of an intervention, and what future change may occur.[276] For example, one study used data from Sicily to create realistic simulations of how people are recruited into organized crime and then tested the effects of implementing various policies to stop that happening (e.g., targeting leaders, providing education support).[277]

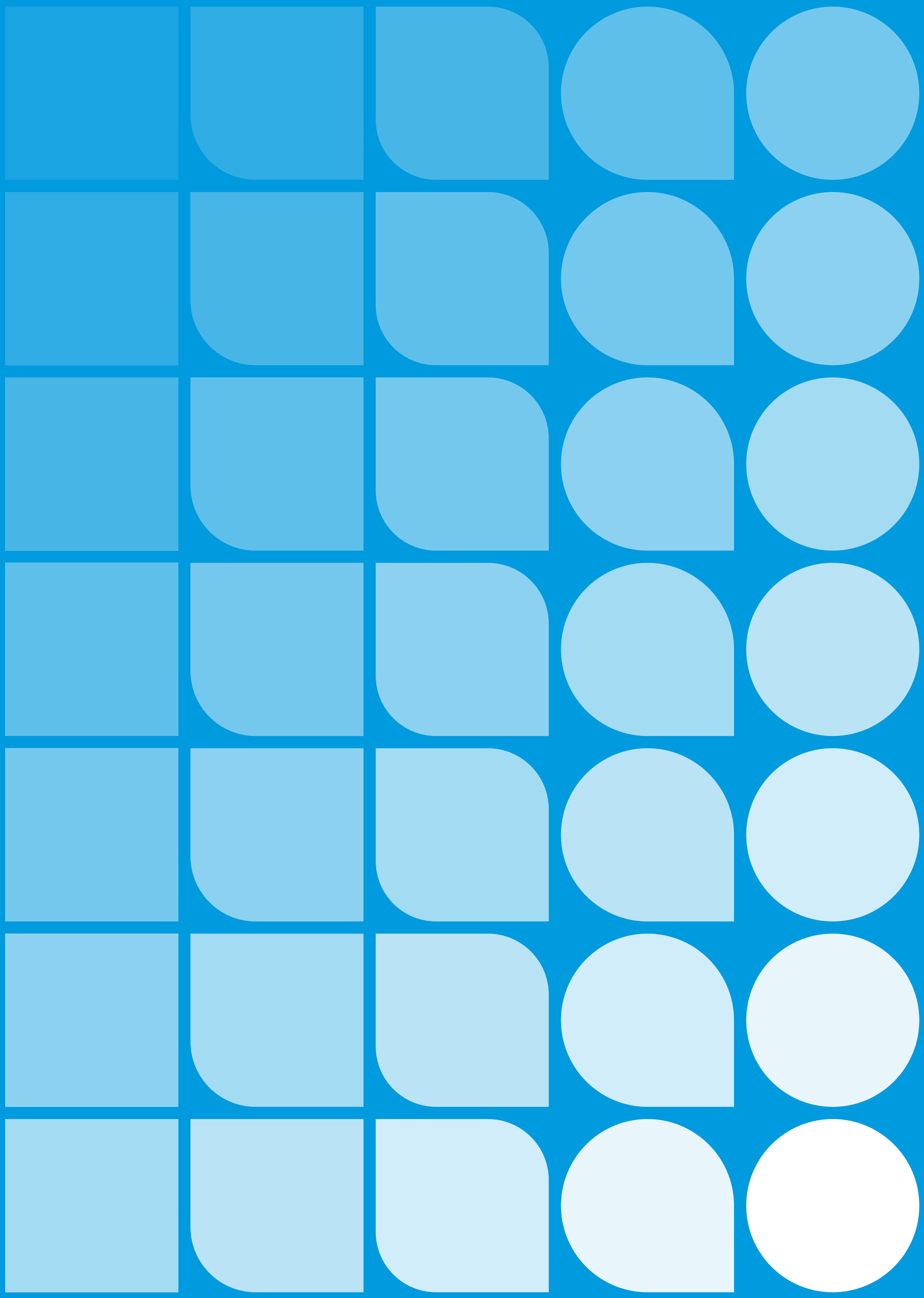
We single out ABM because we think behavioral science can improve it.[278] As we mentioned earlier in relation to macroeconomic models, the agents in ABM are mostly assumed to be operating on rational choice principles.[279] Therefore, there is a big opportunity to build in more evidence about the drivers of behavior - for example, habits and social comparisons.[280]

However, it may not be clear which of these drivers is most relevant for the issue at hand. Therefore, the best course may be to 'include different theories into the model to assess the sensitivity of the results to different assumptions about human decision making'.[281] Doing this allows you to assess how resilient your intervention is to the uncertainties about behavior generated by complex adaptive systems. In other words, an improved ABM could offer a sophisticated way of stress testing policies in advance.

While behavioral science can improve ABM, the requirements imposed by formal models also reveal that behavioral science needs to sharpen up its theories in terms of clarifying causal relationships and how change occurs over time.[282] We build on this point later in our proposal Beyond lists of biases.

PUT RCTS IN THEIR PLACE

One argument is that 'while an initial focus on rigorous empirical research helped BI teams establish themselves in policy making, strict adherence may represent a risk to their long-term growth and relevance'.[283] We agree in the limited sense that finding new options is a good idea: behavioral science should adopt and improve alternative approaches to evaluation. But we stress that the main point is to match the evaluation approach to the type of system and issue at hand. In simpler situations, standard RCTs can work well. When dealing with complex adaptive systems, RCTs can be strengthened to address some of the limitations that become apparent - and the act of doing so should drive behavioral science itself forwards.



05

REPLICATION, VARIATION AND ADAPTATION

> The "replication crisis" of the last decade has seen intense debate and concern about the reliability of behavioral science findings. Poor research practices were definitely a major cause of the replication crisis; the good news is that many have improved as a result.

We need to secure and build on these advances, so we move towards a future where meta-analyses of high-quality studies (some deliberate replications) are used to identify the most reliable interventions, develop an accurate sense of the likely size of their effects, and avoid the weaker options. We have a responsibility to discard ideas if solid evidence now shows they are shaky, and to offer a realistic view of what behavioral science can accomplish.

That responsibility also requires us to have a hard conversation about heterogeneity in results: the complexity of human behavior creates so much statistical "noise" that it's often hard to detect consistent signals and patterns. This much heterogeneity makes the idea of replication itself problematic: a "failed" replication may not show that a finding was false, but rather how it exists under some conditions and not others.

These challenges mean that applied behavioral scientists need to set a much higher bar for claiming that an effect holds true across many unspecified settings.[284] There is a growing sense that interventions should be talked about as hypotheses that were true in one place, and which may need adapting in order for them to be true elsewhere as well.

We need specific proposals as well as narrative changes. The first concerns data collection: expand studies to include (and thus examine) a wider range of contexts and participants, and gather richer data about them. Coordinated multi-site studies will be needed to collect enough data to explore heterogeneity systematically; "crowdsourced" studies offer particular promise for testing context and methods.

We also need to get better at judging how much an intervention's results were linked to its context - and therefore how much adaptation it may need. Behavioral scientists should use and modify frameworks from implementation science to develop such judgment. Finally, we need to develop and codify the practical skills that successfully adapt interventions to new contexts; expertise in behavioral science should not be seen as simply knowing about concepts and findings.

WHAT DID THE REPLICATION CRISIS DO FOR US?

The "replication crisis" of the last decade has seen intense debate and concern about the reliability of behavioral science findings.[285]

This crisis has brought about many advances, and we need to secure and build on them. But we also need to have a hard conversation about how far the underlying goal of creating replicable, generalizable findings is actually achievable.

Replication is seen as 'a cornerstone of science'.[286] An attempt to replicate a study should obtain similar results, if it maintains important features of the original. Most people consider "similar" to mean the effects are in the same direction and are statistically significant.[287] The problem is that large-scale studies have found that only between a third and two thirds of replication attempts produce similar results to the originals - and often the effect sizes are much smaller.[288]

A high-profile example is a 2012 study that found that “signing at the top” of a car insurance declaration increased honest reporting of mileage more than signing at the bottom, as is traditional.[289] The idea was that signing beforehand made ethical concepts prominent to an individual just as they had the opportunity to be dishonest. After BIT failed to find a similar effect in large real-world studies,[290] other careful attempts (including one co-authored by the original researchers) have come to the same conclusion.[291] It could be that this intervention does have an effect in some settings but, on a practical level, there is a real cost to keep testing it rather than other ideas.

Poor research practices were a major cause of the replication crisis. Some of the main ones include publication bias, low statistical power, hypothesizing after results are known, over-reliance on null hypothesis testing, and overly flexible approaches to data collection, measurement tools, and analysis (e.g., stopping data collection once a desired result has been achieved).[292]

The good news is this exposure has improved academic research practices.[293] There are sharper incentives to pre-register analysis plans, greater expectations that data and code will be freely shared, and wider acceptance of post-publication review of findings.[294]

However, we do need to push back against some of the ideas that have emerged from the replication crisis. One is to interpret replication attempts in a binary way - as showing something that “did” or “did not” replicate. There’s a good case that this way of thinking is ‘inadequate’.[295] Instead, replications should be seen as contributions to a larger ongoing meta-analysis, and judged in terms of how much they add to existing stock of knowledge.[296] In fact, this kind of analysis may show that even a non-significant result can advance the conclusion that an effect is real.[297]

This way of thinking points us towards a future where meta-analyses of high-quality studies (some deliberate replications) are used to identify the most reliable interventions, develop a realistic sense of the likely size of their effects, and avoid the weaker options. Operating in the real world incurs real costs; we have a responsibility to discard ideas if solid evidence now shows they are shaky. Overall, rather than over-inflating claims, practitioners should ‘offer a balanced and nuanced view of the promise of behavioral science’.[298]

This vision seems achievable. Evidence suggests that at least some experts in behavioral science do indeed update their views in response to new replication results.[299] More and more reviews based on better quality studies are emerging, giving us a stronger sense of what impact to expect from different kinds of interventions.[300]

IS REPLICABLE, GENERALIZABLE KNOWLEDGE A REALISTIC GOAL?

We may need to rethink our ambitions, however. It’s important to note that there has been a general aim for findings not only to be replicable, but also generalizable. In basic terms, if replicability is about whether the same results would emerge again in the same conditions, generalizability is about whether the same results would emerge again in different conditions.[301]

We usually place greater value on evidence that is generalizable. From a scientific viewpoint, ‘effects that can only be reproduced in the laboratory or under only very specific and contextually sensitive conditions may ultimately be of little genuine scientific interest.’.[302] From a practical viewpoint, we often look for interventions that can be widely “scaled”, and get concerned that effectiveness often seems to fall when they are taken to new places (the so-called “voltage drop”).[303]

Now we can pose the deeper challenge that has been emerging. Is replicable and generalizable knowledge actually a feasible goal for applied behavioral science? Recently, it’s become clear that the complexity of human behavior creates so much statistical “noise” that it’s hard to detect consistent signals and patterns. In other words, results from studies of behavior are highly heterogeneous.[304]

Several large studies have confirmed that results seem to vary between experiments in ways that are difficult to explain.[305] It seems increasingly clear that 'heterogeneity cannot be avoided in psychological research - even if every effort is taken to eliminate it'.[306]

This situation makes replicable, generalizable findings much less likely. One massive analysis found so much heterogeneity across 8,000 studies that it concluded that 'there will always be a sizable chance of unsuccessful replication, no matter how well conducted or how large any of the studies are'.[307] In fact, the idea of replication itself becomes more problematic: a "failed" replication may not show that a finding was false, but rather how it exists under some conditions and not others.[308]

How can we start cutting a path through this thicket? The first step is to carve out two different (but related) drivers of heterogeneity: 1. ***Contexts*** influence results; 2. ***Effects*** vary between groups.[309]

CONTEXTS INFLUENCE RESULTS

Just now, when explaining replication, we said that similar results should emerge if another study 'maintains important features of the original'. But do we know how many and what features are important?[310] When people set up experiments, they make myriad choices about things like the precise wording of messages, the time of day they are given, how participants are recruited, and so on.[311]

Often they are guided by a desire to maximize their chances of detecting an effect.[312]

The point is that these choices vary greatly between studies and experimenters, in ways that often go unnoticed.[313] That's true even for replication studies that explicitly aim to be standardized.[314]

These choices then interact with the inherent complexity of human behavior. The result is that neglected or undocumented aspects of the setting or intervention design may be influencing the observed outcome. When the context changes, those elements may not be present, and so a study may seem to have failed to "replicate" or "scale up".[315]

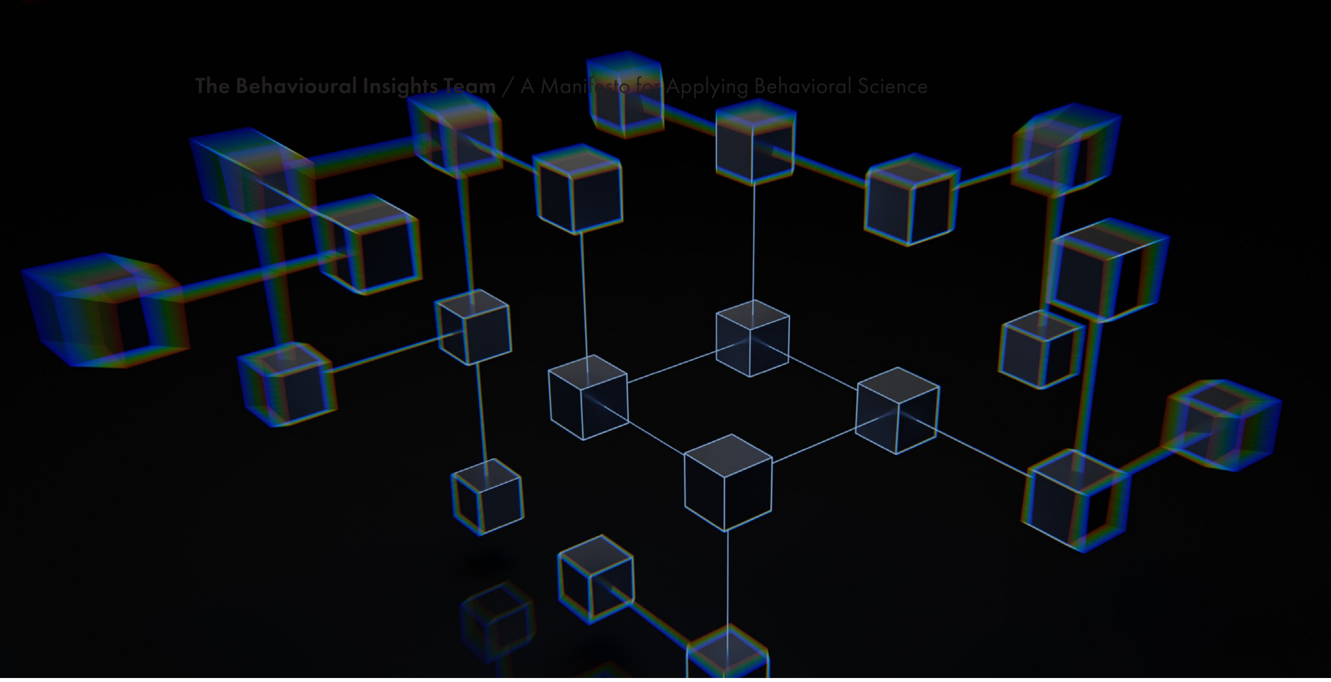
A recent study actually ran an experiment to measure the impact of these contextual factors. Participants were randomly allocated to studies designed by different research teams to test the same hypothesis. For four of the five research questions, studies actually produced effects in opposing directions. These 'radically dispersed' results indicate that 'idiosyncratic choices in stimulus design have a very large effect on observed results'.[316]

In a way, we shouldn't be too surprised. The strong influence of context on behavior has been one of the central insights from social psychology over the last century.[317] Moreover, we are trying to have an effect in the real world. We are not dealing with how laboratories set up experiments. We are dealing with complex adaptive systems that expose people to many cues, and where acting can change the context in unexpected ways.[318] For example, while reminders may influence behavior effectively in initial studies, they may lose their power if people become surrounded by too many reminders, each coming from different sources.[319]

We are also dealing with behavior embedded in institutions and cultures. Those of us trying to shape and improve public policy are confronted by 'intricate bundles of rules, procedures, institutional designs, and contextual constraints'.[320] General concepts coming from studies may only provide only a vague guide here. Similarly, different cultural meanings can render a study ineffective for unanticipated reasons. While the color red is associated with danger in the West, in Chinese culture it symbolizes good fortune. But even cultural variations are not certain to have an effect - they may remain dormant unless 'the right stimulus is triggered'.[321]

EFFECTS VARY BETWEEN GROUPS

Context cannot explain all the heterogeneity in studies of behavior.[322] Another factor is that the effect of an intervention may vary greatly between groups within a population - even though we often talk in terms of an overall "average treatment effect".[323]



For example, a study in India used a smartphone app to give drivers feedback on their driving, using three different kinds of notifications: a reminder of personal best driving performance, average driving performance, or their performance on their latest trip.[324] The first two nudges were effective at improving driving among participants overall, compared to a 'no nudge' control group.

But the study also found significant variation: the "personal best" nudge was significantly better for high-performance drivers who did not seek feedback often; the "average" nudge worked best for low-performance drivers who sought feedback often. Lower performers appeared to be more motivated by goals that were easier to achieve. The study concluded that firms could improve driver performance by 11% through tailored messaging.

Results like these have sparked calls for a 'heterogeneity revolution' in behavioral sciences, which accepts that most effects vary, and rejects the idea that an effect needs to hold across all groups in order to be important or real.[325] Such a call is a challenge to a nudge approach that often prioritizes achieving overall marginal shifts in the behavior of large populations.

This is not just an academic debate. Focusing on the population as a whole may lead to a discussion about whether an intervention "works" that is heavily weighted towards the largest group in the population. An intervention that has a benefit on average may disguise how some groups experience no benefit - or even harms.

Let's link this idea back to the replication crisis again. Effects vary between groups. When an intervention is applied more widely, the makeup of the participants may be different from the original study. That means that the intervention may seem to no longer have an overall effect, even though it is having the same impact on some groups as before.

What looks like a failure to replicate a main finding may actually be driven by more nuanced underlying patterns of behavior that sometimes map poorly against the "haphazard" samples that are collected.[326] An intervention may be dismissed completely when it actually "works" for a meaningful group of people.

For example, the company OPower showed in several studies that energy consumption was reduced by 2%, on average, if customers were shown how their usage compared with that of their neighbors.[327] But later studies showed that the average treatment effect was much smaller when implemented at scale.[328] This is because the intervention had more effect on people with higher incomes and more concern for the environment, and there were more of such people in the initial studies.

MAKE MORE CONSERVATIVE CLAIMS

These challenges mean that applied behavioral scientists need to set a much higher bar for claiming that an effect holds true across many unspecified settings.[329] Our narrative and ambitions may need to change. Over the past decade, many organizations have offered frameworks that aim to raise the base level of behavioral science knowledge (for BIT, these include EAST and MINDSPACE). They offer deliberately simplified messages like "people are influenced by the behavior of others".

We don't need to ban these kinds of statements.[330] But we do need to talk about them more like the way behavioral scientists have talked about incentives: that incentives definitely influence behavior, but they can also backfire, and we can find ways of making them more effective.[331]

We need to be more conservative in explaining that findings may have emerged from a particular context (which is different from saying they are poor-quality).[332] And, we need to be realistic about the size of the impact that interventions based on these findings are likely to have.

It's going to be harder to move away from the simple idea that an intervention "works" - not least because average treatment effects will continue to matter in environments where evidence is scarce and the ability to target interventions is limited. Instead, there is a growing sense that interventions should be talked about as hypotheses that were true in one place, and which may need adapting in order for them to be true elsewhere as well.[333]

We need specific proposals as well as narrative changes. To address these issues, we propose: multi-site studies to explore heterogeneity systematically; better identification of context-dependence in results; and, most importantly, developing skills for adapting ideas to settings.

RUN MULTI-SITE STUDIES TO GET BETTER DATA

Once we understand that results vary by context and groups, the next step is to understand how and why.[334]

First, we can recruit participants using existing methods, but gather richer data about them. To date, only a small minority of behavioral studies have provided enough information to show how effects vary.[335] Moreover, such gaps in data coverage may result from and create systemic issues in society: certain groups may be excluded or may have their data recorded differently from others.[336] Part of the solution may link back to the need for greater diversity among behavioral scientists, since 'diverse researchers collect diverse data' - for example using multiple race-related items, rather than a single binary variable.[337]

Second, we can expand studies to include (and thus examine) a wider range of contexts and participants. Experiments would be constructed to answer for whom the effect appears, and under what circumstances, rather than "Did it work?"[338]

Large replication projects took a step in this direction, but have not been designed to systematically vary aspects of method and context to measure their effects.[339] BIT is attempting to do this for an intervention that uses the "teachable moment" of hospital attendance to reduce future violence.[340]

Here's an analogy. Econometricians carry out "sensitivity analyses" to understand under what conditions their models predict behavior - i.e., stress testing them. Behavioral scientists have minimized the need for these analyses by running experiments - but this means that the specific context of the experiment is fixed, so we are advocating for stress testing findings by varying the context.[341]

'Crowdsourced' studies may be particularly suited to testing context and methods. As explained above, different research teams each come up with a way of testing the same hypothesis, and participants from the same pool are randomly assigned to each. No team sees how the others are approaching the problem. The study can then run a meta-analysis that combines effects across the studies to determine which hypotheses are supported. The ensuing results are the ones that survive the stress test of being subjected to varying approaches (method heterogeneity). A recent study of this kind, with 15,000 participants, found that two of the five hypotheses tested were confirmed by this approach.[342]

In terms of measuring how effects vary by groups, an obvious first step is to recruit more diverse samples. The lack of cultural diversity in participants in behavioral science research has been much discussed, but things are starting to change.[343] For example, the rise of online platforms has made it easier to collect data from participants around the world. A recent study tested the main principles of prospect theory with 4,000+ participants across 19 countries (although most were Western).[344]

Nevertheless, the truth is that collecting varied data better is going to be hard. Organizations like BIT, which work across many different continents, seem to have the greatest ability to deliberately explore how effects vary by group characteristics. Yet there are many barriers to coordinating data collection across different clients and countries, since we are reliant on budgets and willingness to cooperate. Often we need to work with public administrative datasets, but these are often limited (containing just age, gender, and location, for example).

Realistically, practitioners are going to be looking to academia to initiate these advances. And a major investment in research infrastructure will be needed to set up standing panels of participants, coordinate between institutions, and reduce barriers to data collection and transfer.[345] There are promising signs. Recently, projects like Behavior Change for Good have been producing 'mega studies' that test a wide range of interventions with hundreds of thousands of people across many sites.[346] These studies represent a real advance, and the model could be adapted to vary context, methods, and participants even more systematically.

PREDICT CONTEXT-DEPENDENCE

We need to get better at judging how much an intervention's results were linked to its context - and therefore how much adaptation it may need. There are reasons to be optimistic here. One study got people to rate how contextually sensitive a topic was; topics with higher scores were less likely to have successful replication studies.[347] In other words, judgments of context-dependence did seem to link to replication outcomes.

Stronger evidence comes from the 'crowd-sourced' study mentioned earlier. It found that scientists were able to anticipate the results of an intervention just by examining the materials (e.g., sample size, methodology, materials). Most importantly, they could distinguish between studies that were testing the same hypothesis. That means they had a 'fine-grained sensitivity' to which of the design choices would be most effective at realizing the concept in that setting.[348]

The need is to cultivate and teach this sensitivity, so we can predict which interventions are likely to require adaptation, and focus resources accordingly. Currently we don't know how these abilities are developed: behavioral scientist have not reached consensus on which aspects of context matter.[349] BIT and others have sketched out the aspects of 'scalability', but these are still quite basic and focus more on whether the implementation mechanics can scale, rather than if the results will.[350]

Behavioral scientists need to realize that fields like implementation science have thought more deeply about this topic. For example, the Consolidated Framework For Implementation Research (CFIR) was created in order to map the factors involved with moving an intervention from one context to another.[351] These factors include the criteria of "adaptability" and "complexity" and "trialability", which could be used to make the kinds of judgments we outline above. "Adaptability" is based on the idea that there is a distinction between an intervention's "core components" and its "adaptable periphery"; the latter can be modified to fit the setting without undermining the effectiveness of an intervention. Behavioral scientists should be using and modifying frameworks like this to judge how much resource an intervention will need for adaptation.

ADAPT INTERVENTIONS SKILLFULLY

We need to develop and codify the practical skills that successfully adapt interventions to new contexts. The first step is to value them properly. Expertise in behavioral science is still mainly seen as knowing about concepts and findings. That is necessary but not sufficient, given the power of context. We can't simply pluck ideas from a generic "nudge store" and expect them to work.[352] Expertise should explicitly include the ability to understand context and how to make choices that are sensitive to that context. (Our later proposal of how to explore the context fully is relevant here.)

In a way, all we are doing is foregrounding a skill that is at the heart of behavioral science already. Whenever someone creates an intervention, they give an abstract concept concrete form to affect behavior.[353] This act of creation requires many design choices. Which word should be used here? What time of day should the offer be made? How?

We know these choices can make or break an intervention, yet we just don't pay enough attention to them. Many people will have had the experience of being mildly surprised at coming across a study showing a null result - before they see the actual intervention and realize it was put together badly. For example, a 2001 study found no effect of a descriptive norm message (i.e., "most people pay their taxes") on tax compliance. But the message undermined itself by starting with the statement "many Minnesotans believe other people routinely cheat on their taxes"![354]

So far, there seems to be an element of skill or craft to making these choices well - they are not completely reducible to standard procedures, and may never be.[355] For example, standard practice when testing messages is to change the least amount possible between treatment arms.[356] This is a good general rule. But experience shows that it can produce phrases that have a strange and unnatural quality that itself will bias the results. ("If you contact us now, you won't have to worry about this anymore." / "If you don't contact us now, you will have to worry about this more.")

The skill here is knowing how to navigate conflicting priorities and concepts to produce something that works for the particular situation. So, the task of adapting an intervention is actually quite similar to that of creating one in the first place - meaning, once we realize that's the case, we can build on those existing skills.

We also need to go further and adopt a more structured approach to understand and develop this skill. The CFIR framework is useful here as well, since it sets out five major factors affecting how interventions are implemented, including features of the intervention, different aspects of the setting, features of the individuals involved, and so on. Even a basic guide to these points would be helpful, but it's worth noting that they would only be a starting point. The CIFR authors themselves admit that often the distinction between "core" and "periphery" elements 'can only be discerned through trial and error over time'.[357]

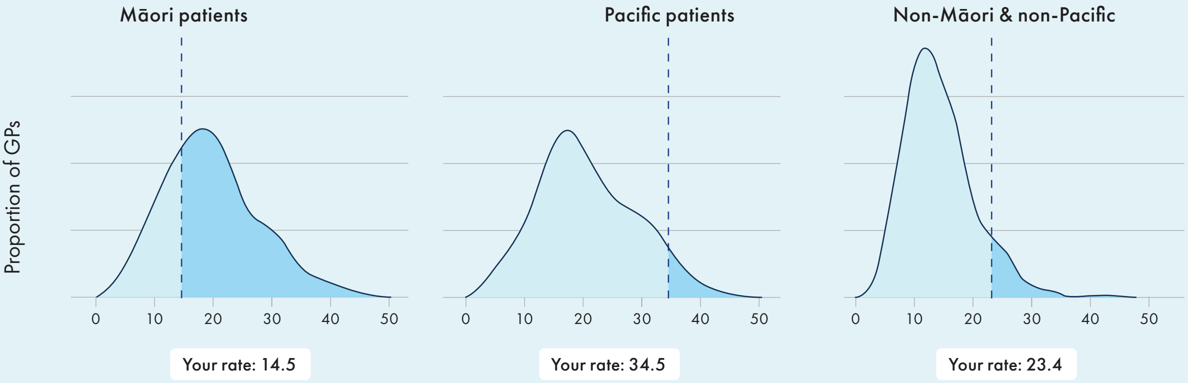
This point means that it's particularly valuable to learn from practitioners how they adapted specific interventions to new contexts. These accounts are starting to emerge, but they are still rare (see BIT example below).[358] Researchers are incentivized to claim universality for their results, rather than report and value contextual details.[359] And of course, these skills can be built through running multiple rounds of prototyping and testing with affected service users and communities.[360]

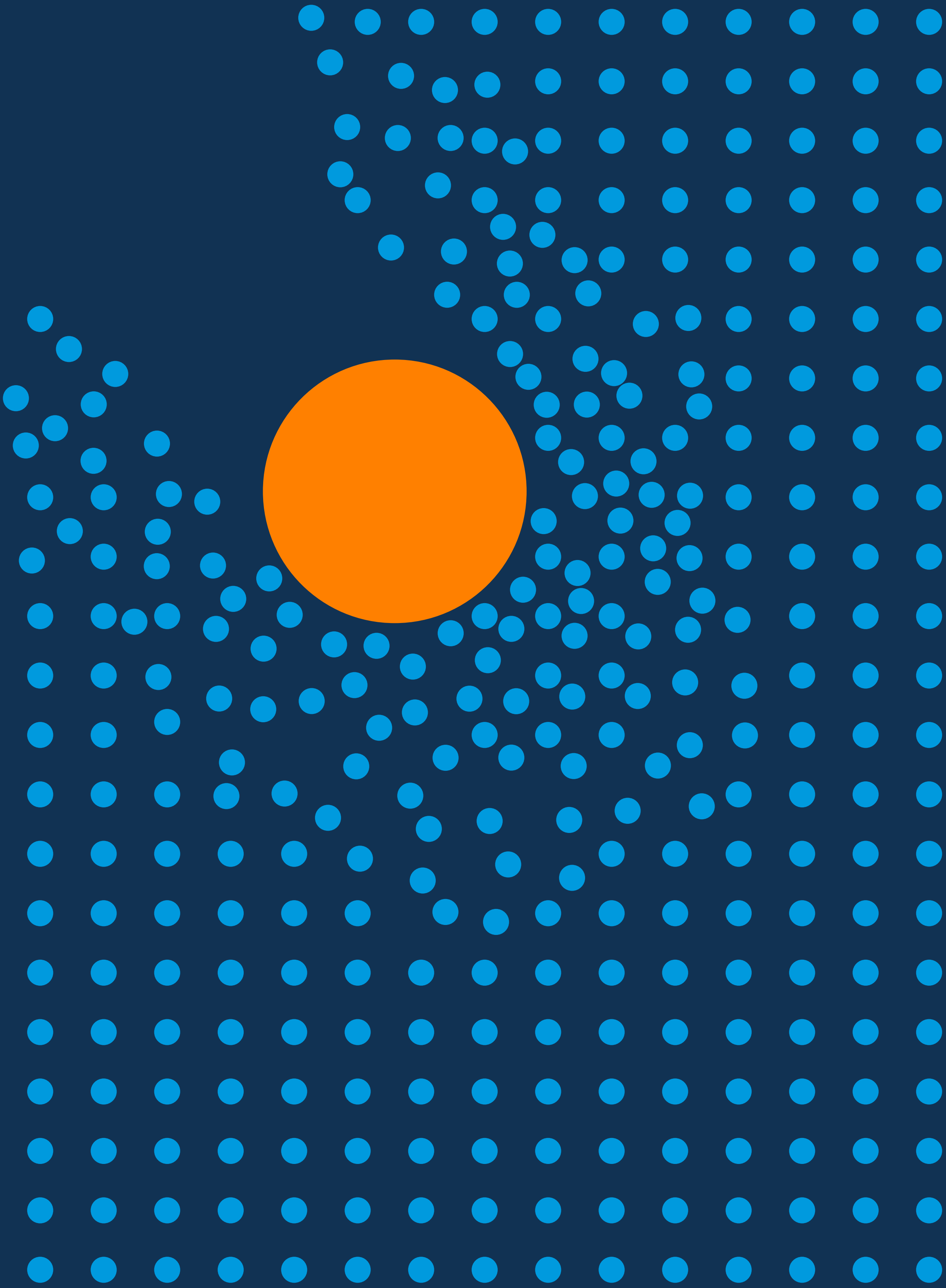
In 2014 BIT worked with Public Health England to create a letter with a social norm message that reduced antibiotic prescribing among high-prescribing primary care providers.[361] *Similar interventions have since been tried successfully in countries around the world.*[362] *When BIT came to adapt the intervention to New Zealand, we knew that Māori or Pacific populations experienced specific challenges and were at a higher risk of infectious diseases.*[363] *They might need more antibiotics, rather than fewer. So we adapted the intervention to show antibiotic prescribing split out by Māori, Pacific, and all other populations. The letter significantly reduced prescribing overall; and we saw no significant change in the prescribing of doctors who were high prescribers overall, but low prescribers for Māori and Pacific populations. Therefore, it appears that health inequities were reduced.*

There is one obvious recommendation missing here. Hopes have emerged that data science techniques will allow us to better understand how effects vary between groups. We agree, but also think that this idea raises big ethical questions. In the penultimate proposal, we will explore the promise and perils involved.

YOUR ANTIBIOTIC PRESCRIBING TO SPECIFIC DEMOGRAPHIC GROUPS*

We know Māori and Pacific patients may need more antibiotics than other New Zealanders. Below is your prescribing rate to different demographic groups, and for GPs in your DHB.





06 MOVE BEYOND LISTS OF BIASES

> Recently there have been claims that our lack of good explanations for heterogeneity is part of a "theory crisis" in psychology. One concern is that psychological theories have become very high-level and vague through their attempts to capture complex behaviors. Another is that theories are specific, but disconnected from each other - and from a deeper, general framework that can provide broader explanations.

This second criticism is very relevant to the way that behavioral science relies on lists of heuristics and biases. These ideas are incredibly useful, but have often been presented as lists of standalone curiosities, in a way that is incoherent, reductive, deadening, breeds overconfidence, and distracts us from answering more important underlying questions (like when does one bias or another apply).

The concern for behavioral science is that it uses and popularizes both these high-level frameworks, like dual process theories, and a collection of disconnected heuristics and biases - with little in the middle to draw both levels together.

We think the way forward is to emphasize theories that are practical. By this we mean: they fill the gap between day-to-day working hypotheses and systematic attempts to find universal underlying explanations; they are based on data; they can generate testable hypotheses; they specify the conditions under which a prediction applies or does not; and they are geared towards realistic adaptation by practitioners.

We think that resource rationality is a good example of the kind of practical theory that applied behavioral science could pursue.

Behaviors often take place in complex systems; intervention effects vary across contexts and groups. Therefore maybe it's not a surprise that we lack good explanations for why findings vary so much.[364] When we fail to obtain an expected finding, we may say that "context matters", but often cannot explain why.[365]

This failure is also an opportunity, since seeking to understand the causes of heterogeneity should lead us to better theories.[366] For example, when data from one of the large replication studies was reanalyzed, it revealed that political ideology was having an important but unidentified effect on heuristics like anchoring.[367]

The need for better theories can be seen as part of a wider "theory crisis" in psychology.[368] We see two different concerns here: one, that theories are too vague and high-level; the other, that they are specific, but disconnected and ungrounded. The concern for behavioral science is that it uses both these high-level frameworks, like dual process theories, and jumbled collections of heuristics and biases - with little in the middle to draw both levels together.

THEORIES THAT ARE TOO VAGUE AND HIGH-LEVEL

Theories of behavior often try to explain phenomena that are complex and wide-ranging.[369] If you are trying to show how emotion and cognition interact (for example), this involves many causes and interactions. Trying to cover this variability can produce descriptions of relationships and definitions of constructs that are abstract and imprecise.[370]

The results are theories that are "weak," in the sense that they are vague and therefore can be used to generate many different hypotheses - some of which may actually contradict each other.[371] The theory becomes hard to test or reject because its looseness means it can accommodate many different findings.[372] And its vagueness prevents us from making useful models or predictions of the conditions under which interventions ought to work (or not).[373] This may have been what happened in the crowdsourcing study mentioned earlier - teams got varying results because the hypotheses allowed many different approaches.

The result is that running experiments does not improve our understanding as it should. We can always claim that a theory is basically true, but there are reasons why it was not supported in this specific instance (e.g., a problem with the way the study was set up).[374] And so weak theories stumble on, unimproved.

We are not saying that this is an accurate summary of how applied behavioral science uses theory. But it does raise questions. Dual process theories arguably provide the main overarching framework.[375] But there are criticisms that they do not sufficiently explain when people use one process or the other, they imply a clear division between the processes that does not exist, they can be hard to refute, and they don't make clear predictions.[376] Models of how and why nudges work tend to be ad hoc and related only to certain domains, making it difficult to predict their effects in new settings - although there has been progress as the field has matured.[377]

THEORIES THAT ARE SPECIFIC BUT DISCONNECTED

On the other hand, we have theories that are more specific, but also disconnected from each other. Some go as far as calling psychology textbooks 'largely a potpourri of disconnected empirical findings'.[378] There can be many different 'mini theories' offering different takes on the same behavior - for example, theories based on cultural factors, emotion or cognition can all be used to explain cooperation in an experiment.[379] Or there may be a precise elaborate theory based on one dataset that only applies in very specific conditions.[380]

At the same time, these theories are often also disconnected from a deeper, general framework that can provide broader explanations (like the theory of natural selection does, for example).[381] If we don't have such a paradigm, we can't predict when we should see an effect or not and we can't distinguish 'results that are unusual and interesting from results that are unusual and probably wrong'.[382] Of course, this disconnect may happen partly because these high-level theories are vague and weak, as noted above.

The main way this issue affects behavioral science is through heuristics and biases.[383]

Examples of individual biases are accessible, popular, and how many people first encounter behavioral science. These ideas are incredibly useful, but have often been presented as lists of standalone curiosities, in a way that is incoherent, reductive, and deadening.

INCOHERENT

The biases usually overlap and conflict in various ways. How are people vulnerable to "optimism bias" - where someone believes that they are less likely to experience a negative event - but also "negativity bias" - where negative things affect our mood and cognition more than positive things?[384] Or how is "optimism bias" meaningfully different from "self-enhancement bias" or "positivity bias"? Where do the boundaries, if any, lie? How do they interact?[385] Presenting lists of biases does not help us distinguish or organize them.[386]

REDUCTIVE

Framing behavioral science as a collection of biases can give the impression that behavior is the product of one factor rather than many. It can create overconfident thinking that targeting a specific bias (in isolation) will achieve a certain outcome - like the illusion of control risk we mentioned earlier in "See The System". And it can give the impression that these biases are universal laws, rather than contingent tendencies.[387]

DEADENING

Repeatedly working from a list in this way may lead to a "painting by numbers" approach that sucks out creativity and innovation from applied behavioral science, leading to repetition. A deeper engagement is more likely to allow unexpected insights into what might work. This risk has not been discussed widely, but it could be increasing as the field matures and broadens.

Perhaps the biggest issue is that looking at lists of biases distracts us from answering the more important underlying questions. When does one or another (conflicting) bias apply, and why? Which are widely applicable and which are highly specific? Meta-analyses of the "paradox of choice" show zero effect overall, but that it can be very powerful in certain settings - what are our theories for why that happens?[388] How does culture or life experience affect whether a bias affects behavior or not?[389] These are highly practical questions when you are faced with tasks like, for example, taking an intervention to new places.

PRAGMATIC RESPONSES ARE NOT ENOUGH TO FULFILL BEHAVIORAL SCIENCE'S POTENTIAL

Given these practical challenges, one response is to step away from theories and just prioritize being pragmatic. Three main approaches in this vein are:

"THROWING THINGS AT THE WALL."

Some people advocate just relying on 'lightweight theorizing and outright guesses', but subjecting them to many rigorous experiments to get reliable findings.[390] You may not have thought too much about what you are testing, but you've got good data on whether it influences behavior.

"POST THEORY SCIENCE."[391]

Another proposal is to use algorithms to uncover new relationships and connections in datasets, and use these findings instead of theorizing. The argument here is that psychology has spent too much time trying to explain the causes of behavior, but can't predict future behaviors accurately.[392] Machine learning can produce insight without theory.

"MUDDLING THROUGH."[393]

In this view, applied behavioral science should be seen as 'a kind of learning by trial-and-error that is informed as much by local, practical knowledge and user feedback as by a universal science of human behavior'.[394] Using behavioral science becomes more like translating some general strategies into specific tactics - and making limited, incremental adjustments that respond closely to contextual needs.

All of these options present problems. For some, costs and inefficiency loom large. Lots of blind incremental experimenting means lots of failure: the risk of the "contextual brittleness" (disconnect between intervention and context) we mentioned earlier is much greater.[395] Having lots of interventions that do not work is fine in contexts where experimentation and measurement is cheap - online environments, for example. But even then there are other costs: an iterative, scattershot approach is likely to take time to produce results.[396] And the truth is that, in many contexts, it does require money to run experiments - so designing your interventions by chance may not seem responsible. Particularly in the public sector, most people expect that someone has thought about what is being offered to them, rather than just sticking a finger in the air.

More generally, rejecting theory means it's harder to learn and make general conclusions. Some have compared this approach to trying to write a novel by collecting sentences from random strings of letters.[397] Even if an algorithm means we are sure that two behaviors are definitely related in a big dataset, we won't know if that relationship holds true elsewhere.[398] We also can't anticipate if an intervention will backfire in a different context.

In other words, giving up on theory for pragmatism will mean that behavioral science cannot meet its challenges or fulfill its potential. For example, being able to tackle bigger, high-level problems may only be possible if you can identify higher-level causes.[399] So what is the best way forward?

BE PRACTICAL

In one sense, the obvious answer is that "we need strong theories, not weak ones". Strong theories are seen as 'coherent and useful conceptual frameworks into which existing knowledge can be integrated'.[400] They identify the most relevant theoretical and empirical questions; try to explain how phenomena do or do not vary across time and contexts; are parsimonious; and can be used to build models that make testable and non-obvious predictions.[401] Various people are trying to propose new ways of developing strong wide-ranging theories,[402] and there are some promising candidates, such as "cultural evolutionary behavioral science".[403]

We could just recommend strong theories as the goal and leave it there. But this definition covers a lot of ground, and we want to focus on helping applied behavioral scientists meet the challenges we face. So, instead we emphasize that a bigger priority for future theories is that they are practical. By this we mean:[404]

1. They fill the gap we've identified in behavioral science: between day-to-day working hypotheses and comprehensive and systematic attempts to find universal underlying explanations. They are not so abstract as to be unhelpful when confronted with a specific issue, but they allow us to raise our eyes above simply understanding what works in the current context.

2. They are based on data rather than being derived from pure theorizing.[405]

3. They can generate testable hypotheses, so they can be disproved.[406]

4. However they also specify the conditions under which a prediction applies or does not.[407]

5. They are geared towards realistic adaptation by practitioners and offer 'actionable steps toward solving a problem that currently exists in a particular context in the real world.'[408] They involve meaningful exchange between those who use them and those who develop them.

Let's be more concrete and give an example of a practical theory that meets these criteria.

A PRACTICAL EXAMPLE: RESOURCE RATIONALITY

Resource rationality is a way of understanding choices and behavior on the basis that people make rational use of their limited cognitive resources.[409] Given there is a cost to thinking, people will only work to find exact solutions if they think doing so is worth the effort. In other words, people look for 'solutions that balance choice quality with effort'.[410]

RETHINKING THE ANCHORING BIAS WITH RESOURCE RATIONALITY

A common example of the anchoring bias involves people estimating a number, but being biased by an initial 'anchor' that puts a specific number in their mind, even if it's irrelevant. For instance, a famous study spun a wheel of fortune in front of participants, who were asked whether they thought the number of African countries in the United Nations was more or less than the number on the wheel. Then they were asked to estimate the exact number. Their guesses were highly influenced by the irrelevant number they had just seen.[411]

The traditional explanation is that people make an initial guess as an anchor, and they try to adjust away from it with more information or thought - but don't adjust enough. While anchoring may seem irrational, a resource rational analysis suggests that people are just trading off the cost of being wrong against the cost of taking time to get the answer more right.[412]

At least one experiment suggests this analysis may be correct. People were asked to predict how long someone will wait at a bus stop, and were given an anchoring time. They were either penalized for how long they took to give a final answer, or for how inaccurate their final answer was. The results showed that people adapted their behavior to get a 'near optimal' tradeoff between speed and accuracy.[413]

Although the basic principle is similar to familiar ideas of "bounded rationality", what's new is that resource-rational analysis offers a systematic framework for building models for how people will act. Proponents call it a 'unifying framework for a wide range of successful models of seemingly unrelated phenomena and cognitive biases.'[414]

More importantly, it's also practical. Take the first criterion for practicality: the need to **fill the gap** between high and low-level theories. Proponents explicitly frame resource rational analysis as 'building bridges' between high-level and lower-level approaches to understanding decisions.[415] (The high-level "computational" one is about working out optimal solutions; the lower-level "algorithmic" one deals with actual cognitive resources and processes.)

This approach is also **based on data** (point 2). It emerged partly from computational analysis and work on 'bounded optimality' in artificial intelligence. Rather than a traditional approach where 'a theorist imagines ways in which different processes might combine to capture behavior', resource rationality focuses instead on evidence of the problems people have to solve and their resources for doing so.[416]

For the three final aspects - **testable hypotheses, specifying conditions, and actionable steps for practitioners** - we look at how resource rational analysis can improve nudging. A recent study on "optimal nudging" starts by echoing some of the criticisms above. The authors also see a high-low level split between 'intuitive, mechanistic accounts of individual nudges' on the one hand and 'formal, abstract accounts of nudging as a whole.'[417] Since models of nudges are often 'domain-specific and ad hoc', we don't have a framework to reliably predict how a nudge will perform for a new context or population. That makes designing nudges time-consuming.

Their solution is to apply the principles of resource rational analysis. Doing so means they see nudges as interventions that change the order or way people encounter information. Nudges make it easier to consider some things rather than others, or change the order in which people think about things. These changes influence people's choices.

The authors used these principles of resource analysis to build models to predict how people would respond to different kinds of nudges, like defaults, and suggested alternatives. They then tested these models against how people behaved in online games where you invest money to earn more, or settle for what you have. When they tested the effect of a default option nudge in such a game, the resource rational model predicted things well: people were more likely to choose the default if the problem was complex, and less likely to if their preferences were different from the average.

The point is that the framework and the models can predict how the effects of nudges will change in new contexts or with tweaks to their presentation. They can also help us understand how effects vary between groups, who may vary in the value they place on certain kinds of information.

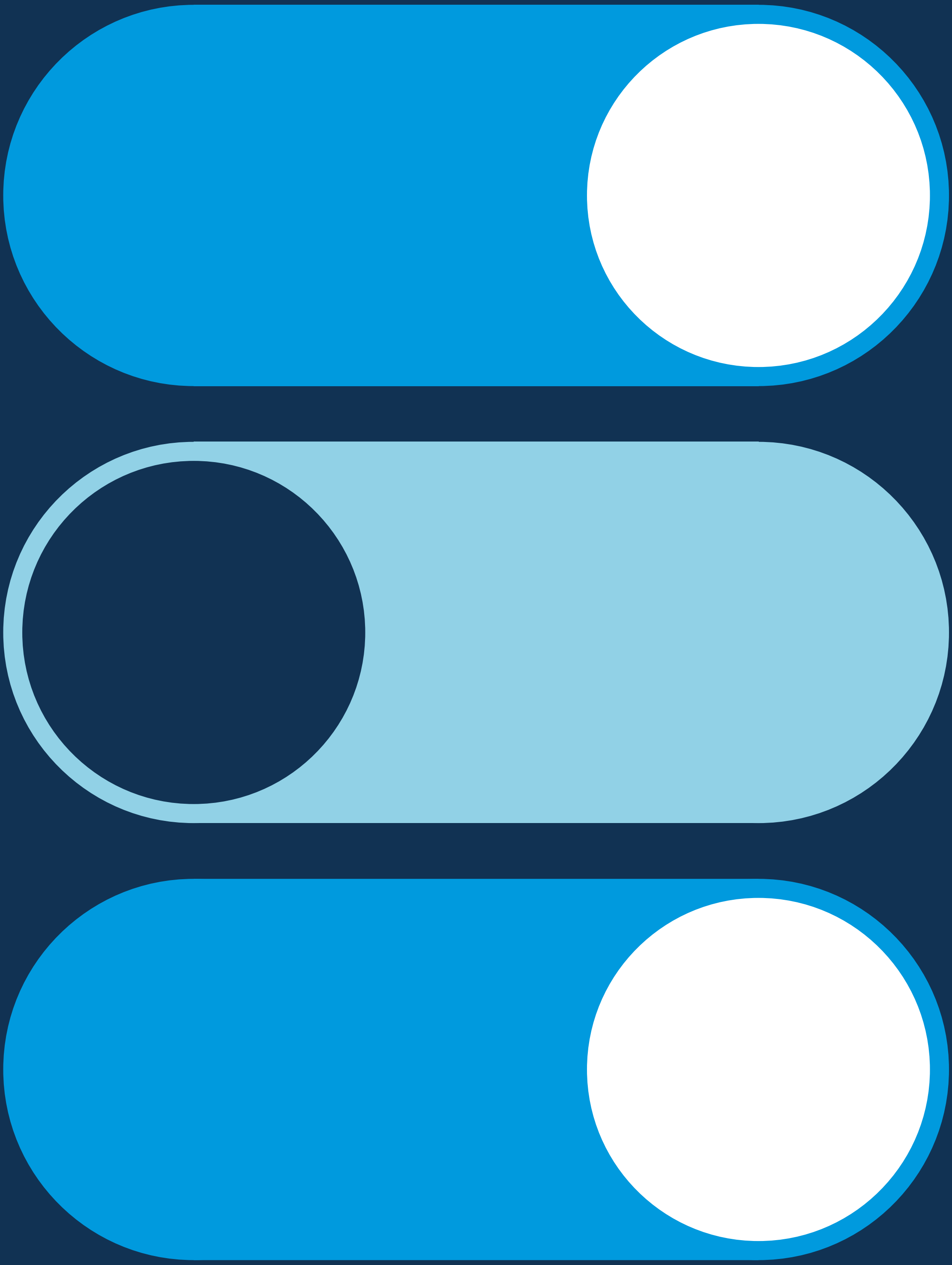
Moreover, the authors built on these results by using algorithms to create 'an automated method for constructing optimal nudges'. When tested against randomly selected nudges, these optimal nudges increased the rewards people received in games and made those choices easier to make (people spent less time spending money to find out information for their decision).

Such an approach could both reveal new kinds of nudges, but also make creating them much more efficient. More reliable ways of developing personalized nudges are also possible, since they would just require some initial data on how an individual makes decisions (recognizing that caution is needed here). These are all highly **practical** benefits coming from applying a particular theory.

THE VISION FOR THE FUTURE

We are not saying that resource rational analysis is completely correct or that it is the only practical theory around. There is still debate about how 'strong' the theory actually is: proponents themselves seem unsure whether it is a unifying framework or just 'a methodological device to efficiently search through the endless space of possible mechanisms.'[418] Many decisions in life are vague and difficult to quantify for input into such a framework, which needs reliable data to work.[419]

Instead, we see resource rationality as having the features we need to aim for in behavioral science theories of the future: a framework that can connect up general and specific ideas; testable, data-driven hypotheses; the ability to make sense of heterogeneity; predictions about when an intervention will work or not; and practical tools informed by the latest data science techniques. That's the vision for moving behavioral science beyond lists of biases and towards greater relevance and sophistication.



07 PREDICT AND ADJUST

> It may seem strange that behavioral scientists are overconfident, since they run so many experiments that may give unexpected results. The reason is hindsight bias - what happens when people feel "I knew it all along", even if they did not. When the results of an experiment come in, hindsight bias may mean that behavioral scientists are more likely to think that they had predicted them, or quickly find ways of explaining why they occurred.

Hindsight bias is a big problem because, alongside overconfidence, it impedes learning, dissuades innovation, and prevents us from understanding what is truly unexpected.

In response, we should develop the practice of getting behavioral scientists to predict the results of experiments, and then feeding back the results to them. Doing so would provide clear, regular performance feedback that is lacking and which enables hindsight bias. But barriers lie in the way. People may not welcome the ensuing challenge to their self-image; predicting may seem like one thing too many on the to-do list; the benefits lie in the future.

We propose: make predicting easy by incorporating it into standard processes; minimize threats to predictors' self-image, for example by making and feeding back predictions anonymously; give concrete prompts for learning and reflection, in order to disrupt the move from surprise to hindsight bias; and build learning from prediction within and between institutions.

Experiments may appear like an admission of uncertainty: we run experiments because we are not sure of something (the notion of "equipoise"). Overconfident people should not run experiments. Yet, as we will explain below, the behavioral science approach can generate overconfidence. So why does experimentation not do more to address this overconfidence?

First, although experimenters are meant to create hypotheses that are confirmed or disconfirmed, most hypotheses are reported to be supported.[420] Moreover, hypotheses may not be particularly detailed - they may not state the expected effect sizes of different treatments, or their likely ranking. This may be particularly true if the study is being conducted by practitioners who have to make tactical decisions, rather than run a structured scientific inquiry.

But perhaps the main reason is one that is familiar to behavioral scientists: hindsight bias. Hindsight bias is 'the belief that an event is more predictable after it becomes known than it was before it became known.'[421] Or, put it another way, it's what happens when people feel "I knew it all along", even if they did not. So, when the results of an experiment come in, hindsight bias may mean that behavioral scientists are more likely to think that they had predicted them, or quickly find ways of explaining why they occurred.

Some aspects of hindsight bias are particularly relevant to behavioral science. One is 'knowledge updating', which refers to how new information is integrated into existing memory. Our memory systems are skilled at fitting new facts into existing schemes - we search out knowledge that is consistent with the novel finding, and thereby make it fit and "feel right". [422] A similar but more sophisticated approach is 'sense-making', which involves the ease with which we can tell a causal story that explains the new information.[423]

Both knowledge updating and sense-making are more likely if a result is not what we expected, since we may start to look for ways to preserve a positive self-image of expertise - "ah, this is in line with this fact I already knew". The key is whether our initial surprise turns into "resultant surprise" and reassessment of views, or whether we can quickly resolve it into a coherent but incorrect explanation that creates hindsight bias.[424]

What are the effects of hindsight bias? For one, it creates general overconfidence in one's predictive abilities, partly because it impedes accurate feedback on them.[425] It also impedes learning, since it can create an illusion of understanding that means we fail to learn from the past.[426] And the feeling that you "knew it all along" may also dissuade you from finding new approaches to a problem, thereby reducing innovation.[427] Finally, hindsight bias can stop us from understanding what is actually unexpected - we do not judge new findings correctly, which stops behavioral science from advancing as a field.[428]

In response, we should develop the practice of getting behavioral scientists to predict the results of experiments, and then feeding back the results to them.

Hindsight bias can flourish if we do not systematically capture expectations or "priors" about what the results of a study will be - in other words, it is not easy to check or remember the state of knowledge before an experiment.[429] Making predictions provides regular, clear feedback of the kind that is more likely to trigger surprise and reassessment, rather than hindsight bias.[430] It could also force people to consider that a range of different outcomes are possible, which can reduce hindsight bias.[431]

Presenting predictions back to participants, along with the actual results, would then provide the clear, regular performance feedback that is currently absent and which allows hindsight bias to develop. Note that the predictions do not need to relate to future studies - the results just need to be unknown to the predictors. And indeed, the predictions may not relate to a controlled study at all, but rather to real-world behavior. But there do need to be clear and specific outcomes to measure predictions against.

The process of predicting and gaining feedback can improve the "calibration" of behavioral scientists - in other words, it can ensure they have appropriate confidence in their judgments. BIT itself has promoted this idea, showing that the judgments of senior officials (and its own staff) can be overconfident, but also that calibration can improve over time.[432]

Prediction brings benefits beyond reducing overconfidence. As noted above, it can create a more accurate sense of how unexpected a set of findings are. Is wider reflection needed? Do beliefs about concepts need updating? Predictions can also reveal the importance of unanticipated null results, thereby reducing publication bias and file-drawer effects.

Regular prediction would also show how expert opinion varies, and whether there are consistently good performers whose views should carry more weight. A recent paper took this further by determining that experts predicted that a video would increase parental uptake of free educational resources by 14 percentage points. In fact there was no effect of the intervention, but the researchers could also measure whether the result was significantly different from the expert prediction (it was). As a result, they could show that their results were worth paying attention to.[433] If we can build up a high-quality set of predictions and predictors, then this will mean better policy advice in situations where tests simply are not possible.[434]

Behavioral science has seen some limited use of prediction as a tool. Results so far are mixed on the predictive powers of behavioral scientists, but the following themes seem to be emerging.

- Experts (and, to an extent, non-experts) can predict the effect of real-world interventions fairly accurately.[435] But, in general, these predictions tend to overestimate the size of effects created by nudge-type interventions.[436]
- Predictions of experts may be more accurate than non-experts - although not always.[437]
- Familiarity matters: practitioners were more accurate than academics at predicting the effect of real-world nudge interventions;[438] non-experts who were likely to be familiar with an intervention were as good as experts at predicting its effects.[439]

We can't be sure of these conclusions because only a tiny fraction of current work involves prediction and feedback. Several barriers lie in the way. People may not welcome the challenge to their self-image that evidence of poor predictions may bring. Real-world trials can involve a frantic process of dealing with practical and conceptual challenges; making predictions can seem like one more, inessential, thing for the to-do list. Even if people understand the benefits, they lie in the future.

These barriers mean that making predictions needs to be supported by institutions and processes. But setting up this support means navigating difficult questions like: What exactly to predict? How to do that? By whom? When? How should results be communicated? And how should participation be incentivized, if at all?

We offer these proposals as a first step towards getting good answers to these questions.

MAKE PREDICTING EASY BY INCORPORATING IT INTO STANDARD PROCESSES

This proposal is obvious, in line with behavioral science principles, and hard to achieve in practice. Organizations need to create an accessible template for collecting forecasts that makes them easy to both send and complete. The template should at least include how to predict outcomes (e.g., estimated treatment effects, direction of treatment effects, significant ranking of trial arms), and whether or how to record one's level of uncertainty.

A prompt to access this template should be inserted at a point after interventions are near-final, but before a trial is launched - so predictions can influence which interventions are selected, or any that need to be improved. (It's worth noting, though, that if people know predictions will influence an outcome, that may distort their predictions!) Ideally, a specific project member or role should be clearly responsible for ensuring predictions have been made. That same person should then trigger the individual results feedback, once findings are confirmed (perhaps through an automated mail merge).

MINIMIZE THREATS TO PREDICTORS' SELF-IMAGE

Being presented with evidence of inaccurate predictions may threaten a predictor's self-image of expertise, prompting them to engage in hindsight bias as a way of reasserting their agency.[440] This tendency could be minimized by presenting all predictions as reasonable "based on what you knew then". Such a formulation presents the results as a challenge to one's past self, rather than one's present or intrinsic abilities. Updating of beliefs may be easier to achieve as a result.

Predictions should be made and fed back anonymously, since this further minimizes perceived threats to self-image and encourages re-calibration.[441] Note that predictions would be known generally (to aid better group calibration), just not who made them. However, there are clear advantages to identifying a group's most accurate predictors, since their views should be given greater weight when deciding which interventions to test - and potential acclaim could act as an incentive for participating.[442]

GIVE CONCRETE PROMPTS FOR LEARNING AND REFLECTION

Perhaps the biggest priority is to ensure that feedback leads to learning and adjusting. Understanding the 'epistemic emotions' created by prediction feedback is key here. Common epistemic emotions are surprise, confusion or curiosity, and they are triggered by conflict between existing and new knowledge.[443] Surprise is particularly important, since it can be used as a prompt for learning and reorganizing networks of knowledge.[444] However, what can happen is that 'initial surprise' turns into hindsight bias, instead of the 'resultant surprise' that can lead to reorganizing knowledge.

The practical implication is to challenge and disrupt the move from surprise to hindsight bias. An obvious starting point is how the results are framed - prompts could make it easier for the recipient to reassess and update their knowledge. A very simple one might be, "If this is true, what else do you need to reassess?" Another idea is to ask predictors to include a reason for their prediction at the time they make it. Presenting this past reasoning alongside the results may disrupt the move to coming up with new explanations in the present.

BUILD LEARNING WITHIN AND BETWEEN INSTITUTIONS

This kind of adjustment and learning is much easier to achieve in conditions of psychological safety, where people and groups can freely say that they got things wrong and accepted ideas can be questioned. BIT is currently working to set up its own formal forecasting and score-keeping system - and we are aware that incentives and leadership will be key to ensuring a prediction culture takes root. Ideally, there would be a wider commitment to learning, so that the organization builds a central store of predictions and findings. Such a store could also exist between institutions, as shown by the example of the Social Science Prediction Platform (www.socialscienceprediction.org). This idea reinforces our earlier calls for behaviorally-enabled organizations.

VALUES

08 BE HUMBLE, EXPLORE AND ENABLE

> Behavioral scientists may over-confidently rely on decontextualized principles that do not match the real-world setting for a behavior. Deeper inquiry can reveal reasonable explanations for what seem to be biased behaviors. Therefore, we should: avoid using the term "irrationality", which can limit attempts to understand actions in context; acknowledge that our diagnoses of behavior are provisional and incomplete ("epistemic humility"); and design processes and institutions to counteract overconfidence.

When exploring behaviors, we need to: pay greater attention to people's goals and their own interpretations of their beliefs, feelings, and behaviors; reach a wider range of experiences, including marginalized voices and communities; and recognize how apparently universal cognitive processes are shaped by specific contexts, thereby unlocking new ways for behavioral science to engage with values and culture.

In addition, more can and should be done to broaden ownership of behavioral science approaches. One route is to enable people to become more involved in designing interventions themselves - and "nudge plus", "self nudges", and "boosts" have been proposed as ways of doing this. But these new approaches should not be seen simplistically as "enabling" alternatives to "disempowering" nudges. We propose a new set of criteria for deciding when enabling approaches may be appropriate: opportunity; ability and motivation; preferences; learning and setup costs; equity; and effectiveness.

A final piece missing from current thinking is that enabling people can lead to a major decentering of the use of behavioral science. A range of people could be enabled to create nudges for positive societal change (with no "central" actors involved). Rather than just creating self-nudges through altering their immediate environments, they may decide that wider system changes are needed.

Behavioral scientists who succeed in the future will be ones that are humble about their own knowledge, who carefully explore people's intentions and experiences, and who enable individuals to use behavioral science to achieve personal and collective goals.

BE HUMBLE

We need to recognize the limits to our knowledge as behavioral scientists. Instead, several factors can push us towards being overconfident in our judgments and the likely consequences of our actions.

Of course, it's important to note that overconfidence affects experts in a wide range of fields.[445] In our Behavioral Government report, we showed how and why policy makers are overconfident in their predictions about people's behavior. And overconfidence is a particular risk when dealing with complex adaptive systems, since we may misidentify causes and effects by applying simpler mental models that produce illusions of control.[446]

Nevertheless, overconfidence can emerge because of reasons that are specific to behavioral science.

Behavioral science may be seen as providing a technical justification for seeing decisions as flawed, and thus in need of corrective action. A good example is the behavioral economic concept of 'internalities', which involve a struggle between a person's present self (who may want a cake) and their future self (who desires to be healthy). In practice, policy makers often choose to prioritize the future self, which often loses out otherwise.[447] These behavioral concepts, and the language of bias and irrationality, can end up increasing the confidence of choice architects in terms of what people "really" or "should" want, and why they are acting as they are.[448]

In reality, it may be that people are acting on the basis of reasons that a choice architect has not perceived or anticipated. People may have specific contextual knowledge that cake is unlikely to be available in the near future, for example. Indeed, new evidence from 169 countries shows that, as economic environments worsen, 'there is a stronger and more consistent tendency to discount future values'.[449] A recent study showed that apparently irrational loss-gain framing effects may actually reflect an awareness that others may punish you for ignoring such effects.[450] The ready technical explanation offered by behavioral science could provide confidence that obscures the need to search more deeply for less obvious explanations.

These technical explanations may also be based on 'an overly cognitive view of people as individual decision-making agents, rather than as social humans who are embedded in established practices and networks'.[451] Questions of culture and society, which are likely to complicate matters, may be downplayed as a result.[452]

These factors mean that behavioral scientists may over-confidently rely on decontextualized principles that do not match the real-world setting for a behavior. Such 'contextual brittleness' is likely to lead to poor outcomes.[453]

Again, there are many instances where behavioral scientists have avoided the issues just outlined. BIT has always emphasized the importance of understanding behaviors in context and in depth, going beyond ostensible explanations:[454]

ENERGY SWITCHING

Many households do not switch energy suppliers, despite the potential for large savings; this inertia may seem particularly "irrational" for lower-income households. But during our work on this topic, we came across stories that made the inaction more reasonable. Over time, lower-income households (in particular) may have learned the collections process of their current company. They know that a "final warning" is not final, and they still have two more weeks before they lose power. Effectively, they can then use the power company as a line of credit to prioritize other payments. But they lose this valuable knowledge if they switch providers. Apparent inertia may actually be a considered strategy.

HEALTHCARE CHOICES

Our previous work on reducing patient waiting times in the UK had identified that primary care doctors were generally sending patients to the same set of secondary care providers.[455] The issue is that these referrals happened even when those providers had no capacity. Patients would experience long waits, even though there were nearby alternatives with good availability. In many cases, these choices were as suboptimal as they seemed. But deeper exploration showed that sometimes other factors were at play. Unsurprisingly, patient transport was an important factor. Larger, better-known hospitals tended to attract more public transport routes than the alternative providers. So, although these alternatives were all within a few miles of the patient, in some cases they might be less accessible.

ROAD USE

Going into our work to improve the health and safety of food delivery agents in Australia's gig economy, we knew that many workers were frequently breaking road rules. In particular, they rode bicycles on footpaths and other pedestrian-only areas. This was happening despite strong disincentives: there was a significant risk of incurring fines that could wipe out a whole day's earnings (and social media posts from workers suggested that they were well aware of this risk). However, observing the behaviors in context revealed that the footpaths often ran next to narrow, busy roads with heavy truck and bus traffic. So, the agents were trading off the risk of a fine against the risk of injury or death on the roads - a different calculation from the one we had perceived initially.

How can we make this kind of deeper inquiry more likely? Here are some proposals:

AVOID THE TERM "IRRATIONALITY"

We will not focus on the long-running and complex debates about how to define rationality and irrationality.[456] Our main concern is with the consequences of how 'irrationality' is used in practice. The act of diagnosing irrationality in others seems to imply that you yourself are rational, or at least have the ability to detect rationality. At the same time, the act can delegitimize the views of the 'irrational' party - their disagreement is not valid because they are not playing by the rules of reason, unlike you.[457] Doing so can lead to failure to understand the reasonableness of people's actions. Of course, we need to avoid setting up straw men representations of how behavioral scientists think, as some critics do. But we think the use of the label 'irrational' may increase over-confidence and impede deeper inquiry. Dropping it may be a necessary, but not sufficient, way of solving these problems for practitioners.[458]

EMBRACE EPISTEMIC HUMILITY

Epistemic humility is based on 'the realization that our knowledge is always provisional and incomplete - and that it might require revision in light of new evidence.'[459] For behavioral scientists, this might involve recognizing what initial inquiries are essential, rather than simply reaching for a familiar theory, concept or intervention and applying it to the situation at hand. It might be about pausing to reflect on how far existing knowledge can be transferred between contexts and domains. It might be about recognizing the possibility of backfires and spillovers, of recognizing how goals and preferences (including our own) may be complex and ambiguous. The Covid-19 pandemic has shown just how difficult it can be to predict behavior, and could act as a spur for recognizing why greater epistemic humility is needed.[460]

DESIGN PROCESSES AND INSTITUTIONS TO COUNTERACT OVERCONFIDENCE

While behavioral scientists should be familiar with the concept of overconfidence and its causes, applying such ideas to our own actions is much harder. Instead, we should look at how to design and redesign the contexts in which behavioral scientists are making decisions in order to promote greater humility. In its Behavioral Government report, BIT has already shown how policy makers are often affected by issues such as optimism bias and illusions of control.[461] We then set out a series of institutional changes to reduce their ensuing overconfidence, including pre-mortems, reference class forecasting, and break points. What are the equivalent changes to processes that might reduce overconfidence among those applying behavioral science?

A common theme through these ideas is the need for more and better inquiry into behaviors in context, rather than making assumptions. At BIT, we refer to these inquiries as Explore work.

EXPLORE

Open-ended qualitative exploration of the context and drivers for behaviors is not new to the behavioral sciences.[462] BIT itself has always emphasized the importance of investing in this kind of inquiry - and has worked to incorporate new methods of doing so, such as citizens' juries.[463] As part of this effort, BIT has released a new Explore report, an accessible guide for exploring behaviors in order to create interventions. However, three areas demand particular focus in the future.

The first is to pay greater attention to a) people's needs and their existing strategies to fulfill their needs; and b) people's own interpretations of their beliefs, feelings, and behaviors. Of course, there is existing work to build on here. Design thinking offers ways of understanding users' needs, how they may be met, and how they may reshape the nature of the "problem" itself.[464] So-called "wise interventions" foreground 'the meanings and inferences people draw about themselves, other people, or a situation they are in', and try to reshape these interpretations to change behavior.[465] These approaches have been less prominent (but not absent) because the last decade has seen more focus on changing contexts and situations than attitudes and intentions.[466]

Second, we need to explore a wide range of experiences, including making greater efforts to reach and recognize marginalized voices and communities.[467] We need to understand how structural inequalities can lead to expectations and experiences varying greatly by group and geography.[468] We need to be aware of exactly whose experience we are exploring and what may be shaping that experience.

We may need to collect more nuanced data on, say, race and ethnicity, rather than relying on the limitations of administrative data.[469] And, finally, we need to acknowledge that the behavioral scientists doing this work bring their own perspectives, given the field itself is still relatively homogeneous - as we discuss in more detail below.

Finally, we need to see how better exploration can reveal the interplay of context, culture, and cognition. Some make the case that recent applications of behavioral science have focused mainly on universal cognitive processes, triggered by immediate cues, as drivers of behavior; less attention has been paid to the role of culture, society, and values.[470] Taking this focus has pros and cons.[471] One risk is that it neglects the wider significance of explore work, instead seeing it as serving only a limited set of functions - e.g., constructing tailored interventions for specific circumstances, or checking how fixed concepts (like "anchoring") play out in practice.

In reality, a growing body of research shows how context, culture, and cognition are interrelated, and that neglecting one limits your understanding of the others.[472] As one summary puts it, 'universal cognitive processes are shaped by the specific cultural repertoires provided by the social environment, which vary between cross-cutting social groups.'[473] For example, memory can function differently in Western and East Asian cultures because of varying conceptions of time.[474]

In other words, there is dynamic interplay between the kinds of things that Explore work uncovers on the one hand, and behavioral science concepts on the other. The former is not somehow secondary to the latter. People's experiences of culture and group identity profoundly influence the way that social norms function, rather than being just interesting variations on a central concept that remains unaffected.[475]

Explore work - done well - can reveal how these interactions take place. In doing so, it can unlock new ways for behavioral science to engage with values and culture, addressing the criticism that the approach has 'a thin conception of the social'.[476] For example, one influential view of culture is that it influences action 'not by providing the ultimate values toward which action is oriented but by shaping a repertoire or "tool-kit" of habits, skills, and styles'.[477] There are similarities here to the heuristics and biases 'toolkit' perspective on behavior, and you can see how behavioral scientists could start explaining how and when certain parts of the toolkit become more salient to people.

Since this proposal may seem abstract, a final example may be useful. "Scarcity" is a well-known behavioral science concept. As formulated by Mullainathan and Shafir, scarcity explains that people with a limited resource (e.g., money, food, time) will focus narrowly (or "tunnel") on managing that resource.[478] However, tunneling their cognitive energy in one direction (e.g., providing food today) means they end up neglecting other concerns, such as longer-term planning. Scarcity then becomes a driver keeping people in poverty, as well as one of its effects.

- Scarcity is a powerful and useful idea that has created valuable research and policy initiatives.[479] However, it has been criticized along the following lines, which echo the arguments we make above:[480]
- The scarcity model assumes that those with limited resources find economic need to be the most salient form of scarcity.

- However, they may be faced with multiple forms of scarcity at once (e.g., health issues, economic insecurity, violence).

- When trying to prioritize, 'hierarchies of concern are far from universal; culturally available frames influence which of several competing needs takes priority to become the object of "tunneling"'.[481]

- In other words, what it is appropriate to "tunnel in" on may vary according to groups. People with few means may spend their limited money on funerals because worries about the social cost of not doing so outweighs narrow economic concerns.

- By failing to consider the way that values influence the operation of scarcity, research 'risks labeling behaviors that enable survival in specific contexts as "'non-optimal"'.'[482] As we noted above, applying a general concept universally may lead us to see certain choices as irrational - wrongly.

Careful Explore work can avoid these problems by identifying the different forms of scarcity present and how people interpret them using items from their cultural 'toolkit'. Doing so would help us to better understand these high-level concepts and identify the conditions when they do or don't influence behavior. And, in turn, this would prevent instances of the 'behavioral brittleness' mentioned earlier, where decontextualized cognitive frameworks are poorly aligned to specific contexts.

ENABLE

Explore work helps intervention designers, but we can also go further. The people affected by an intervention could be involved in designing it themselves; they could also be enabled to drive changes that benefit themselves or society at large.

Going further like this is important because many applications of behavioral science have been top-down, with a 'choice architect' enabling certain outcomes.[483] Critics have argued that such a setup is problematic because it is manipulative, identifies people's preferences poorly, and limits transparency, learning and autonomy.[484]

We do not explore those questions here.[485] Instead, we emphasize that BIT and others have always supported and practiced a wider range of empowering approaches. From the start, the field has promoted techniques such as using natural frequencies to improve understanding of statistics, an idea that is often placed in opposition to nudging. This fact is only surprising if you have a simplistic either/or view of nudge versus other approaches; the reality is that a broader palette is and always has been in use. The way this palette should be used is still under debate - there are also hurdles to making deliberative engagement feasible in practice.[486]

As we explain below, more can and should be done to broaden ownership of behavioral science approaches. Attention has focused on three main ways of doing so:

NUDGE PLUS

Nudge plus is where a prompt to encourage reflection is built into the design and delivery of a nudge (or occurs close to it). People cannot avoid being made aware of the nudge and its purpose, enabling them to decide whether they approve of it or not. While some standard nudges, like commitment devices, already contain an element of self-reflection, a nudge plus must include an 'active trigger' - just having the potential for reflection 'is not sufficient to prompt deliberation and cause lasting behavior change'.[487]

SELF-NUDGES

A self-nudge is where someone designs a nudge to influence their own behavior. In other words, they 'structure their own decision environments' to make an outcome they desire more likely.[488] An example might be creating a reminder to store snacks in less obvious and accessible places after they are bought.

BOOSTS

Boosts emerge from the perspective that many of the heuristics we use to navigate our lives are useful and can be taught. A boost is when someone is helped to develop a skill, based on behavioral science, that will allow them to exercise their own agency and achieve their goals.[489] Boosts aim at building people's competences to influence their own behavior, whereas nudges try to alter the surrounding context and leave such competences unchanged. The similar but lesser-known idea of 'steer' preceded boosts.[490]

When these ideas are discussed, there is often an underlying sense of "we need to move away from nudging and towards these approaches". But to frame things this way neglects two crucial questions: how empowerment actually happens; and when to use the different approaches.

EMPOWERMENT CUTS ACROSS EXISTING LABELS

Right now, there is often a simplistic division between disempowering nudges on one side and enabling nudge plus/self-nudges/boosts on the other. In fact, these labels disguise two real drivers of empowerment that cut across the categories. They are:

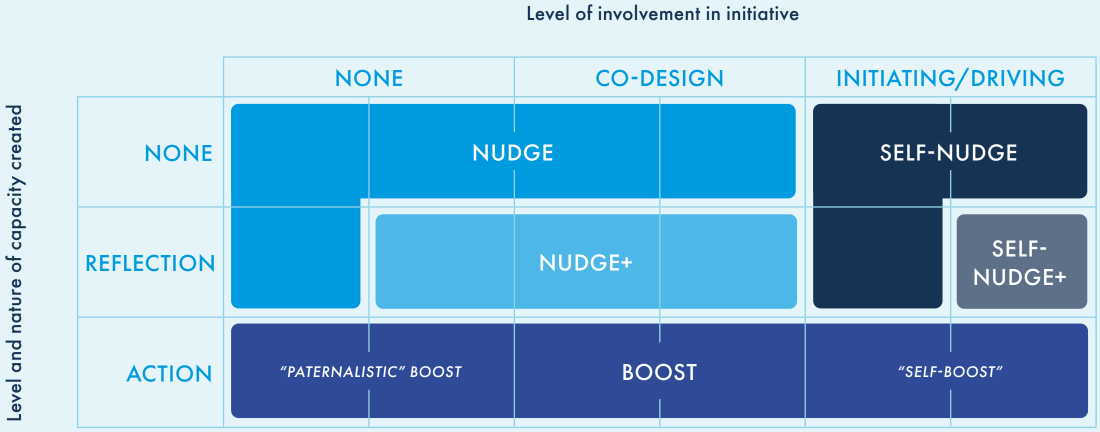
1. How far the person performing the behavior is involved in shaping the initiative itself. They could not be involved at all, involved in co-designing the intervention, or initiating and driving the intervention itself.

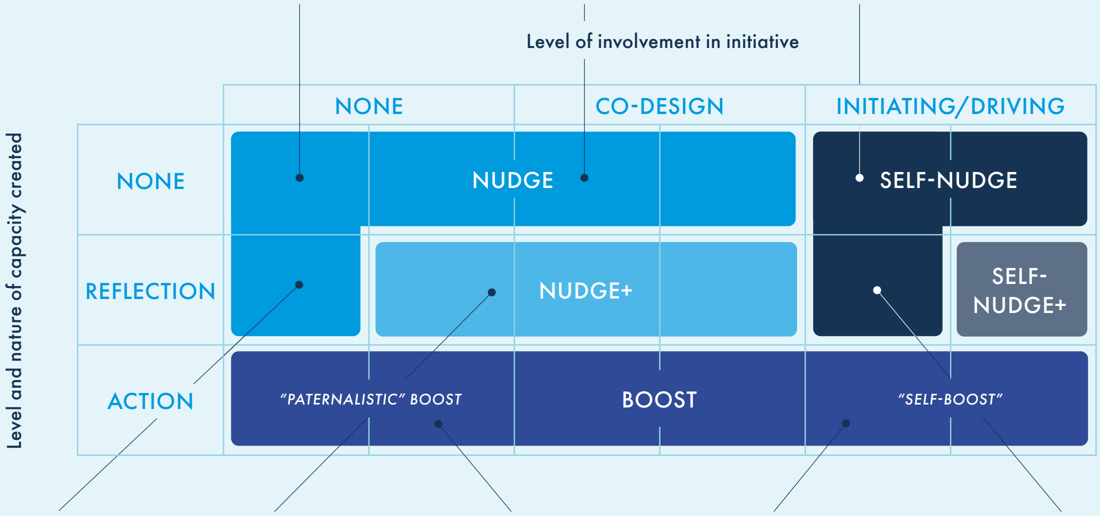
2. The level and nature of any capacity created by the intervention. The intervention may create none (i.e., have no cognitive or motivational effects), it may create awareness (i.e., the ability to reflect on what is happening), or it may build the ability to carry out an action (e.g., a skill).

In the figure opposite, we show how the different proposals map against these two drivers.[491]

We now want to highlight the main things this figure reveals. Co-design requires some explanation. Co-design uses creative methods 'to engage citizens, stakeholders and officials in an iterative process to respond to shared problems'.[492] In other words, the people affected by an issue or change are involved as participants, rather than subjects. This involvement is intended to create more effective, tailored, and appropriate interventions that respond to a broader range of evidence.[493]

These co-design methods can mesh well with behavioral science approaches. Both have a pragmatic focus on what can be done in the real world; both highlight aspects of behavior that escape rational actor models.[494] Facilitated co-design sessions could help participants see how findings from behavioral science are relevant to an issue and decide what interventions are justified. For example, the UK's move to automatic enrollment for pensions was preceded by deliberative events, including a National Pensions Day where 1,000 people from across the UK were presented with evidence on the issue; 72% voted to adopt automatic enrollment with the right to opt out (and 20% for no right to opt out).[495]





e.g. healthy foods placed prominently in canteen

e.g. employees deciding they want healthy foods to be prominent & developing arrangement

e.g. creating regular automatic bank transfers to savings account

e.g. gyms introducing commitment devices to increase usage

e.g. users working with gyms to develop commitment devices that actively remind users of their exercise goals

e.g. initiator decides that the best option is to teach people heuristics to increase savings

e.g. individual reads about heuristics and constructs own heuristics for savings goals

e.g. placing letters to be posted on doorhandle

The point here is that people may be heavily engaged in selecting and developing a nudge intervention that nonetheless does not trigger any reflection or build any skills (the main focus of approaches such as nudge plus and boost). People may choose, with consideration and full information, that they do not want those things to happen.[496]

This choice is similar to the idea of self-nudging, in the top right-hand corner of the figure. People are enabled to create their own nudges that they may then forget about, even as they continue to work. They have exercised agency, even though they may not experience autonomy later, while the nudge operates.[497]

In contrast, in the bottom left of the figure, a policy maker has decided that the best option is for someone to be taught a "boost" (e.g., simple rules of thumb for managing finances). In the absence of greater engagement, there is a risk that this becomes "paternalistic boosting", where a policy maker has assumed that people will want this approach. While proponents of boosts say that 'individuals choose to engage or not to engage with a boost', they also say 'sometimes, lack of motivation may even be addressed with specific boosts (or nudges).'[498] These two things seem to be in tension.

"Paternalistic boosting" is in contrast with the bottom right corner of the figure, "self-boosting", where people learn about heuristics and then proactively apply them to their own challenges, absent any prompting from a policy maker.[499]

Our point is that these distinctions about different ways to enable are absent from current debates about "nudging versus boosting".[500] Identifying these two drivers of involvement and capacity show how the way in which these approaches are applied matters.

There is a final element that is missing from the current debate – how far enabling people can lead to a major decentering of the use of behavioral science.

The approaches in the figure above tend to assume either that a central actor is creating the intervention or, if the person concerned is also the creator, then the intervention is focused on themselves.

But if more people are enabled to use behavioral science, they may decide to introduce interventions that influence others. Rather than just creating self-nudges through altering their immediate environments, they may decide that wider system changes are needed instead. In other words, a range of people could be boosted to create nudges that generate positive societal change (with no 'central' actors involved).

These are not new ideas. In 1969, George Miller encouraged psychologists to find how best to 'give psychology away'.[501] In a little-noticed section of Nudge, Thaler and Sunstein actually propose that 'workplaces, corporate boards, universities, religious organizations, clubs, and even families' could decide to nudge themselves.[502] The creators of Nudge Plus do suggest that it could 'provide a link between citizen action on public policy issues and bottom-up movements for social and political action.'[503]

However, we need more examples of how this can be done. One might be the way that the Fair Tax Mark campaign emerged in the UK. The Mark was a sign that a vendor had not engaged in tax avoidance; it acted as a nudge, since it provided a signal to consumers without restricting their choice.[504] But the Mark was created by a not-for-profit actor, rather than by a government. One could imagine that many other groups could be enabled to launch campaigns like these that draw on behavioral science principles.

These ideas point towards a new dynamic for the public sector. Proponents of co-design argue that more traditional, top-down approaches are inadequate for addressing wicked problems, whose multi-dimensional nature requires input from both experts and the public.[505] This idea chimes with our proposals above on 'system stewardship'. If there is an increasing need to shift away from top-down control towards enabling conditions for behavior, then this changes the role of policy designers. Rather than being architects, they may need to be more like facilitators, brokers, and partnership builders.[506]

We think this is the right direction of travel, and it shows how behavioral science can strengthen democratic engagement rather than weaken it.[507] But to move forward we need a clearer understanding of the different ways people can be enabled, rather than trading competing labels.

When to use more enabling approaches?

The second crucial question is not 'is boost or nudge best?', but how to match the approach to the situation.[508] We propose that the following criteria should be used to inform that judgment:[509]

1. OPPORTUNITY

How likely is it that people will be able to use enabling approaches at the moment of action? Put differently, how will people know that this is the moment to "strategically call on automatic processes" if automatic processes are already in play?[510] There is much evidence that being aware of decision-making biases is very different from being able to counter them.[511] Checks are needed as to whether realistic opportunities exist and how they can be exploited (e.g., by creating new habits).

2. ABILITY AND MOTIVATION

Sometimes people's true preference is not to be actively engaged in a choice or behavior (e.g., choosing not to choose). Similarly, 'if individuals lack the cognitive ability or motivation to acquire new skills or competences, then nudging is likely to be the more efficient intervention'.[512]

3. PREFERENCES

If someone's preferences about an issue are difficult to discover, then more enabling approaches may be better than nudging. This is also true if goals vary widely within a population, or if the same person has conflicting goals.[513] Enabling approaches may also be less successful if the target behavior is not aligned to the individual's interests.[514]

4. LEARNING AND SETUP COSTS

The investment required from individuals (learning costs) and policy makers (setup costs) may vary between approaches. While creating boosts may take less effort than traditional education, they 'may require some hours of instruction and practice' and 'a regular investment of time'.[515] In contrast, a nudge 'typically reduces the cognitive and other costs of decisions for choosers, who are freed from devoting time and attention to the problem'.[516]

5. EQUITY

If there are learning costs, who will want to bear them? This matters because, as proponents say, 'only those people who seek the competence offered by a boost will adopt it'.[517] It is likely that the people with the most time and resources will engage and stick with the learning process, which means there could be inequitable outcomes. Equally, approaches that require less buy-in can reduce inequities.[518]

6. EFFECTIVENESS

Finally, there is the question of which approach is likely to be most effective at influencing behavior. Various claims have been made about why nudge plus or boosts are likely to produce deeper and longer-lasting changes than nudges alone.[519] Studies that compare effectiveness are just emerging - some showing nudges outperforming boosts,[520] some the reverse,[521] and others that nudge plus does better than nudge alone.[522] Clearly, the field is still in the empirical foothills of being able to compare different approaches for certain behaviors.

These criteria are all technical ones; autonomy and transparency may be valued as principles that outweigh all other considerations. But we argue that it's important for the Explore phase to consider these criteria, rather than assuming the kind of intervention that people prefer.

09 DATA SCIENCE FOR EQUITY

> Recent years have seen growing interest in using new data science techniques to reliably analyze the heterogeneity of large datasets. The idea is to better understand what works best for certain groups, and thereby tailor an offering to them.

This vision is often presented as straightforward and obviously desirable, but it runs almost immediately into ethical quandaries around manipulation and discrimination. There is also emerging evidence that people often object to personalization.

We propose that the main opportunity is for data science to identify the ways in which an intervention or situation appears to increase inequalities, and reduce them.[523]

We call this idea data science for equity. It uses data science to support not exploit. But it needs to be supported by other attempts to increase agency, and more data on which uses people find acceptable.

THE PROMISE OF MACHINE LEARNING

There are now more datasets available and better techniques to analyze them.[524] While the term "data science" covers a wide range of activities, there has been particular interest in machine learning, a field that develops algorithms that can offer new insights into the heterogeneity we just discussed.[525] Rather than explaining the details of machine learning, we want to focus on the results it can produce.

Investigating how effects vary by group is not a new idea. The idea of "segmentation" has been core to marketing for decades.[526]

Analyzing trial results by subgroups is a common step, albeit one that requires care, as the replication crisis showed.[527] Machine learning offers more sophisticated, reliable and data-driven ways of detecting meaningful patterns in datasets.[528]

- A machine learning approach has been shown to be more effective than conventional segmentation approaches at analyzing patterns of US household energy usage to reduce peak consumption.[529]
- Danish data has shown that machine learning can accurately predict whether someone has a bacterial infection, thereby complementing physicians' attempts to limit unnecessary antibiotic prescribing.[530]
- The smartphone driving app in India mentioned earlier used machine learning to identify how feedback messages matched to driving abilities.

There has been much excitement around the idea that these results could be used to create more effective targeted interventions. In other words, people should receive the type or variation of intervention that works best for them - or they should be proactively targeted based on their predicted needs.[531] "Scaling" a program becomes about mass tailoring rather than uniform deployment.

There are claims that these personalized or "algorithmic" nudges are much more effective than generic ones,[532] particularly so-called "hypernudges" that can create highly-personalized online environments.[533] More generally, the rise of bespoke interventions is often seen as the future for applied behavioral science.[534] Usually the implicit or explicit message is that we should just accelerate towards this future as fast as possible.[535]

Many theories offer reasons why personalization should produce positive results, such as: perceived self-relevance; feeling of rightness or fit; familiarity or fluency; and self-efficacy.[536] Real-world empirical studies also provide support.[537] But we need to recognize that this vision of tailored inventions requires major technical, ethical and acceptability challenges to be overcome.

TECHNICAL CHALLENGES

For a start, some of the same criticisms from psychology's replication crisis may also apply to machine learning. Measurement, model selection, and the way claims are communicated have all been the targets of criticism.[538] A common concern is whether the datasets that "train" machine learning algorithms introduce bias and reduce their generalizability.[539] Some studies also question the added value of machine learning. One looked at whether researchers could predict certain life outcomes of children; complex machine learning models often failed to outperform simple benchmark approaches.[540]

Then there is the effort required, since 'personalizing nudges is not a small endeavor'.[541] While machine learning may produce interesting results, data scientists actually spend more than half their time doing the mundane, painstaking work of cleaning data first.[542] Accessing these datasets often raises privacy concerns, since they may include sensitive personal data.[543]

Finally, the recommendations need to be feasible for policy makers. Choosing the right number of segments involves balancing analytical concerns against knowledge about how many can be practically implemented.[544] And the truth is, at least for the public sector, the infrastructure for delivering tailored interventions at scale often just does not exist. While it may exist in the private sector, this is often in online environments only; physical environments are much harder to personalize.[545]

ETHICAL CHALLENGES

Adopting new technologies is never value-free.[546] There is no "view from nowhere" here either: using machine learning is not just a technocratic task. Ethical quandaries soon emerge.

First, using new techniques means that people are unlikely to know what data has been used to target them, and how.[547] Arguably, individuals who are affected by algorithms are owed an explanation of how a decision has been made - but that can be hard when those algorithms are continually adapting and processing new data.[548] Transparency is particularly important if the effects are widespread and lasting (one study found that targeted digital advertisements altered people's self-perceptions, with effects lasting at least two weeks).[549]

Machine learning datasets may feature detailed and sensitive information about an individual, including their biases and vulnerabilities.[550] When combined with low transparency, the result could be greater opportunities to manipulate recipients into outcomes that are badly aligned with their preferences.[551] If done systematically at large scale, this could even lead to market manipulation.[552] We need to make sure that increased data capabilities do not just result in instruments of precise control (or perceived control).

Closely related to manipulation concerns is the fear that data science will create new opportunities to exploit, rather than help, the vulnerable. One aspect is algorithmic bias. Models that use datasets that reflect historical patterns of discrimination can produce results that reinforce these outcomes.[553] For example, an algorithm will have a biased understanding of what a "successful" candidate is if it uses a data from a company with discriminatory hiring practices. Examples range from racial bias in healthcare provision to gender bias in credit scoring, although some argue that algorithms are actually less likely to be biased than human judgment.[554]

The UK's Department of Work and Pensions (DWP) is trialing an algorithm that analyzes past historical data to predict which cases are fraudulent in the future. The UK's National Audit Office reports that the DWP is aware that 'biased outcomes' could occur, since 'if the model were to disproportionately identify a group with a protected characteristic as more likely to commit fraud, the model could inadvertently obstruct fair access to benefits.'[555]

The DWP is attempting to manage these risks by:

- Pre-launch testing and continuous monitoring.
- "Fairness" analysis, which looks at how false positive results are distributed across groups with protected characteristics.
- Keeping the final decision with a human, and not telling caseworkers why each case has been flagged for review.

Since disadvantaged groups are more likely to be subject to the decisions of algorithms, there's a particular risk that inequalities will be perpetuated.[556]

And even if an algorithm is not making a decision, but just identifying people's situations, that itself can be used to exploit them. For example, new ways of identifying financial security could be used to target support more effectively - or to target high-interest loans more lucratively.

ACCEPTABILITY CHALLENGES

There is also emerging evidence that people object to personalization. While they support some personalized commercial services, like shopping and entertainment, they consistently oppose advertising that is customized based on sensitive information.

Moreover, there is an "acceptability gap": even if people accept a personalized service, they are generally against the collection of sensitive information that personalization often relies on.[557]

While systematic data on acceptability is limited, there are many cases where people have an instinctive negative reaction against personalization.[558] When a company tries personalization that crosses into being "creepy," uproar and damage to their reputation can ensue. The Dutch bank ING found this when it tried to introduce targeted advertising for additional products, based on their customers' behavior: after a massive backlash, the Dutch Data Protection Authority formally warned against this kind of marketing.[559]

Reasons why people react negatively to personalization include: perceiving it as an invasion of privacy, as when the data is too detailed (e.g., a previous transaction history) or clearly comes from another website;[560] if it is seen as an explicit attempt at manipulation;[561] if it contains already-known content;[562] or if it seems to be based on a stereotypical judgment.[563] For example, when overweight people thought they had been given information about a weight loss program based on their weight, rather than randomly, this led to more negative thoughts and lower intention to perform healthy behaviors.[564]

DATA SCIENCE FOR EQUITY

How to navigate these challenges? We propose that the best way forward is for data science to identify the ways in which an intervention or situation appears to increase inequalities, and for behavioral science to be used to reduce them.[565] We call this idea data science for equity: using the power of data science to support not exploit.

Behavioral science can make a particular contribution here. Since humans are training the machine learning models, we need to understand how biases in our perceptions and judgments are transferring to algorithms.[572] There is growing interest in the ways that behavioral scientists can help understand and modify the "behavior" of algorithms, as part of a wider "behavioral data science" agenda.[573]

For example, groups that are particularly likely to, say, miss a filing requirement, could be offered preemptive help. Algorithms can be used to better explain the causes of increased knee pain experienced in disadvantaged communities, thereby giving physicians better information to act on.[566] They could identify when frontline workers are discriminating (consciously or not) against service users.[567] Or they could spot outliers who are achieving better outcomes than their situation would predict, and try to understand why - so good practices could be facilitated more widely.[568]

There are reasons to be optimistic. Biases in algorithmic judgments may be easier to fix than those in human judgment, and increasing effort is going towards this goal. For example, many models that tried to predict postpartum hemorrhage had 'high risk of bias' that could 'perpetuate or even worsen existing disparities in care'.[569] Additional work produced new models that managed to reduce these biases.[570] Best practice guides are now recognizing the need to build user experience perspectives into data science projects, so the effects of algorithms are considered fully.[571]

"Data science for equity" may seem like a platitude, but it's a very real choice: the combination of behavioral and data science is powerful, and has been used to create harm in the past. Nor is it a simple choice: we need to work through the implications of "data science for equity". The idea may be more ethically justified because it tries to reduce harm and injustice and help those who are vulnerable.[574] But this is only one strand of the objections we just considered. Behavioral scientists need to consider all the "who", "how", "when", and "why" questions:

- **Who** does the personalization target, and using what criteria? Many places have laws or norms to ensure equal treatment based on personal characteristics, such as age, race, and gender. When does personalization violate those principles? Does personalization undermine social cohesion by exacerbating existing fault lines?[575] These questions actually go beyond behavioral science, but they should not be ignored.

- **How** is the intervention constructed? To what extent do the recipients have awareness of the personalization, choice over whether it occurs, control over its level or nature, and the opportunity for giving feedback on it?[576]

- **When** is it implemented? Is it at a time when the participant is vulnerable? Would they likely regret it later, if they had time to reflect?

- **Why** is personalization happening? Does it aim to exploit and harm or support and protect the recipient, recognizing that those terms are often contested?

		"WHY": PURPOSE OF PERSONALIZATION	
		EXPLOITATIVE	SUPPORTIVE
"HOW": LEVEL OF AWARENESS, CHOICE, CONTROL, OPPORTUNITY FOR FEEDBACK	LOW	Manipulation that the field needs to reject	"Data science for equity" may be justified in some instances
	HIGH	Will not be tolerated, and therefore unlikely to occur	Long-term goal for the field, but also presents practical barriers

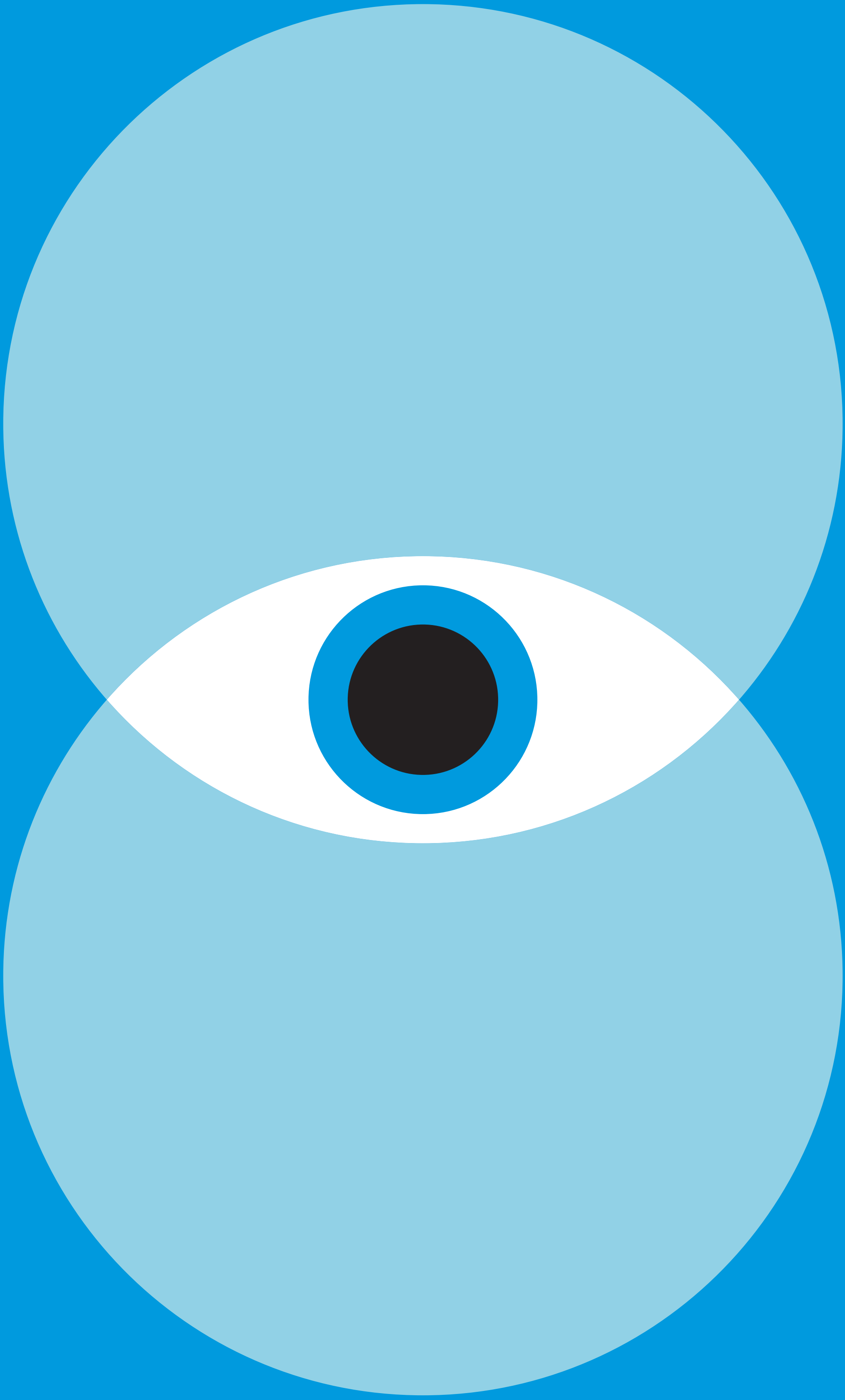
The "how" and "why" are perhaps the two most important factors for the idea of data science for equity. How far does a supportive intent justify an intervention that may not be transparent or apparent? As the simple table below indicates, the goal would be to move the "data science for equity" agenda from the top right-hand corner to the bottom right-hand corner. In other words: aim for a situation that reduces the ethical issues with how data science is being applied as well.

Getting there will take time. But we can identify some obvious priorities now. One is to get more data on how acceptable people find different kinds of personalization to be. What tradeoffs would people find acceptable in order to receive higher value personalized services?

Behavioral science can help guide our questions here as well. For example, one possibility is a move towards using algorithms to personalize prices. Behavioral science offers existing evidence about how people make judgments about the fairness of pricing which can be relevant here.[577] People may tolerate high price tailoring if the products are very similar or firms can explain why the differences are fair (e.g., the higher price cross-subsidizes those who can't afford as much).[578]

A more practical goal is to increase people's agency in the face of personalization, perhaps by developing their ability to detect tailoring or targeting when it happens.

For example, a recent study found that a short intervention that prompted people to reflect on an aspect of their personality increased their ability to detect advertisements targeted at them by up to 26 percentage points.[579] Teaching these kinds of techniques may be a good option, since they could be resilient to constantly changing technologies and strategies.



10 NO 'VIEW FROM NOWHERE'

> Behavioral scientists need to understand how they bring certain assumptions, privileges, and ways of seeing to what they do. We are always situated, embedded, and entangled with ideas and situations. We cannot pretend there is some set-aside position from which to observe the behavior of others - there is no "view from nowhere".

Behavioral scientists may not see the extent to which they hold elite positions that stop them from understanding people who think differently. In addition, homogeneity in terms of gender, race, physical abilities, sexuality, and geography also influences the viewpoints, practices, and theories of behavioral scientists. So, rather than claiming that science is value-free, we need to find realistic ways of acknowledging and improving this reality.

We propose: improved self-scrutiny, with practitioners querying how their identities and experiences contribute to their stance on a topic; new ways for potential subjects of research to judge researchers and decide whether they want to participate; wider participation in intervention design to bring in new perspectives; and greater diversity among behavioral scientists, through actions like professional networks connecting the Global North and Global South.

Exploring, enabling, and decentering - the proposals we made just now raise the question of who is acting. It's a crucial point. Deeper exploration may not be possible if you have not reckoned with where you are starting from. Applying a behavioral lens may be insufficient if you only consider who is seen through the lens, and not who is looking from the other side - or what the act of "seeing" entails.[580]

Our final proposal is one of the most wide-ranging, challenging, and important. Behavioral scientists need to understand how they bring certain assumptions, privileges, and ways of seeing to what they do.[581] We cannot pretend there is some set-aside position from which to observe the behavior of others. We are always situated, embedded, and entangled with ideas and situations.[582] Yet it may not seem this way, since these entanglements constitute our way of seeing itself.

We sum up this idea as 'no "view from nowhere"'. For the philosopher Thomas Nagel, the "view from nowhere" was an objective stance that allows us to 'transcend our particular viewpoint'.[583] We do not think achieving such a stance is possible for behavioral scientists. No objective observation deck outside society exists; the place from which we see behaviors and construct issues matters. Assuming otherwise can create harm.

This section looks first at how the characteristics and positions of behavioral scientists as people can influence both their findings and the methods and concepts they use. We end by trying to offer ways of making behavioral scientists more aware of their perspectives, of diversifying the field of behavioral science, and of introducing greater equality between the people on either side of the behavioral science "lens".

WHO BEHAVIORAL SCIENTISTS ARE INFLUENCES WHAT THEY DO

We start by considering the position of applied behavioral scientists. This is a group usually defined by having knowledge, skills, and education. Many of them are in positions where they can shape actions (public and private) through their advice and findings. Therefore, it's fair to say that behavioral scientists as a group have a fairly privileged or elite position, especially when it comes to policy questions.

This elite positioning means that certain ideas are likely to get strong support in ways that escape full awareness. For example, one critical paper argues that behavioral economists overstate the value of saving for retirement because they are rationalizing their own tendency to save more than average. Similarly, the paper also argues that their professional background means they rarely understand stances that are skeptical of education.[584] We suggest that most people working to increase uptake of the Covid-19 vaccines have been vaccinated themselves.

The issue is not with behavioral scientists holding these positions per se: it's about awareness. We may not see the extent to which we hold elite positions that are preventing us from understanding people who think differently. There is evidence that politicians and bureaucrats misperceive the opinions of both the public in general and the users for whom they design services.[585] In a recent study, US "policy elites" (e.g., judges, media pundits, lobbyists, scientists) inaccurately perceived issue opinions by around 14 percentage points.[586]

Often, such elites believe that others' opinions are more similar to their own than they actually are, and people will act the way they would.[587] In our Behavioral Government report, we called this the "illusion of similarity".[588] The danger is that policy elites are placing their group values and preferences on others, while thinking they are adopting a view from nowhere. This does not mean that policy makers can never act, but rather that they need to carefully understand their own positionality and that of others before acting.

The risk of this happening is not limited to the policy domain. The positioning of funders, researchers, and those researched affects the way many interventions are realized. In response to a local resident asking, "Why am I always being researched?", the US non-profit Chicago Beyond argues that:

'Funders are often cast as "outside of the work," and researchers as
objectively neutral and merely "observing the work." This does not account for the biases and perspectives every person brings to the work. When data is analyzed and meaning is derived from the research, the power dynamic often mutes voices of those who are marginalized.'[589]

Of course, behavioral scientists have other aspects of identity apart from educational and professional status, including gender, race, physical abilities, sexuality and geography. These features also influence our viewpoints.[590] And here again there have been concerns that particular perspectives dominate. Only a quarter of the behavioral insights teams cataloged in a 2020 survey were based in the Global South.[591] An over-reliance on using English in cognitive science has led to the impact of language on thought being under-estimated.[592] In the US, there have been particular concerns over the lack of racial and ethnic diversity in behavioral and social sciences, which are less diverse than biomedical sciences or engineering.[593]

Who behavioral scientists are influences what they do. The positions and perspectives of those studying behavior influences their theories and methods.[594]

So this raises questions about the accuracy and legitimacy of conclusions emerging from a relatively homogeneous elite.[595] In the words of Neil Lewis, Jr., 'Those in dominant positions within society and within disciplines do not notice the centrality of their positions, and as a result end up assuming that their patterns of thoughts, feelings, and behaviors are "normal" and neutral'.[596]

These assumptions then form a base for developing theories, conceptualizing variables, collecting data, and interpreting findings.[597] For example, data from western, educated, industrialized, rich, and democratic (WEIRD) samples may be seen as more generalizable to humans as a whole.

One study examined this possibility by analyzing whether 5,000 psychology articles specified the racial/ethnic/national/cultural characteristics of the sample in question.[598] The authors argue that doing so signals limited generalizability - i.e. "this conclusion only applies to x group".

They find that samples from the United States were featured less in titles compared to both other WEIRD and non-WEIRD regions. At the same time, samples from the US were featured more if they referred to minorities 'who may be perceived as exceptions to assumed generalizability of the White American population'. In other words, there may be an assumption that findings from White US residents are particularly generalizable. Hence there are now calls for psychology to generalize from - rather than just to - Africa.[599]

We can step back and see these arguments as part of a bigger debate about the role of science. The "view from nowhere" is one of positivism: the scientific method is a value-free pursuit of objective truths about human behavior that rise above politics and beliefs. But this perspective has been criticized through a long history of skeptical inquiry. Some have seen the scientific method as entrenching power structures or representing a particular Western perspective, for example.[600]

We are not proposing abandoning the scientific method; we have seen that randomized trials can lead to prior assumptions being questioned and even overturned. But we do agree that 'values are entering the very process of producing knowledge within the behavioral sciences'.[601] Rather than trying to deny this reality in favor of a 'view from nowhere', we should find ways of ensuring it produces better outcomes.

MORE SCRUTINY

We're conscious that we don't have all the answers here. A starting point could be for behavioral scientists to cultivate awareness of how their stances and viewpoints affect their practices. At one level, this task would involve researchers examining their relationship to and feelings about the topic in question, and querying how their identities and experiences contributed to that stance.[602] Hypothesis generation could particularly benefit

from this exercise, since arguably it is closely informed by the researcher's personal priorities and preferences.[603] Working through the reasons why one is pro-vaccination may be the first step to developing the ability to perceive and feel how someone holds an opposing view - perhaps in terms of concerns about purity and trust.[604]

This self-reflexive scrutiny could also be applied to theories and methods.

Behavioral scientists could be actively reflecting on interventions in progress, including what factors are contributing to power dynamics.

'How are inequitable approaches, methods, measures filtering into the study, and what are opportunities to do differently?'[605] The task may be uncomfortable, but attempts to change approaches can succeed. Anthropology moved away from claiming objectivity to recognizing the central role played by the researcher's perspective.[606]

Self-scrutiny may not be enough. We should also find ways for people to judge researchers and decide whether they want to participate in research. We mean this in a broader sense that goes beyond consent forms. For example, Chicago Beyond has created a handbook that provides community organizations with criteria for clarifying what they want to get out of research and whether they want to participate.[607] As we said before, the behavioral lens should work both ways.

MORE PARTICIPATION

If people do decide to participate, that participation can be set up to challenge the assumed 'view from nowhere' - including through co-design approaches discussed earlier. Experiments with politicians have shown that misperceptions regarding public opinion can be reduced by increasing exposure to voters.[608] The nonprofit Ideas42 gives the example of a project in New York City where feedback from participants revealed that a major barrier to using new waste disposal units was the need to touch a handle. The handle was often dirty and people were on the way to work or school. The intervention designers note that while 'our team's missed opportunities to improve key design elements illustrates our own biases', these could be limited through resident participation.[609]

Looking beyond specific interventions, wider participation in behavioral science work could be a way of keeping the field fresh by bringing in new perspectives. Rory Sutherland has emphasized

> how behavior can be successfully influenced through the "alchemy" of asking non-obvious questions and adopting unique perspectives.[610] Applied behavioral science may be weakened if it is a closed system limited to expert input only. We noted earlier the deadening effect of a "painting by numbers" approach that just uses a limited palette of heuristics and biases.

MORE DIVERSITY

If personal viewpoints and situations constrain behavioral science, then expanding the range of people who become behavioral scientists should allow the field to learn more. For example, building on our earlier point about hypothesis creation, there is evidence that cognitive diversity may particularly benefit 'complex, multi-stage, creative problem solving, during problem posing and hypothesis generation'.[611]

Many kinds of diversity could be targeted. In terms of geographic diversity, projects in low- and middle-income countries should have researchers from those countries fully integrated. Doing this requires addressing barriers like a lack of professional networks connecting the Global North and Global South, and the time required to build understanding of the tactics required to write successful grant applications from funders.[612] Within higher-income countries, much more could be done to increase the racial diversity of the behavioral science field, whether in terms of support for starting and completing PhDs,[613] or reducing the significant racial gaps in publicly-funded research that exist in some countries.[614] These changes require sustained effort, and may not happen quickly, but are necessary for the future credibility of the field.[615]

ENDNOTES

1 Hallsworth, M. & Kirkman, E. (2020) Behavioral Insights. MIT Press.

2 Chater, N., & Loewenstein, G. (2023). The i-frame and the s-frame: How focusing on the individual-level solutions has led behavioral public policy astray. Behavioral and Brain Sciences. DOI: https://doi.org/10.1017&am. Schmidt, R. (2022). A model for choice infrastructure: looking beyond choice architecture in Behavioral Public Policy. Behavioural Public Policy, 1-26. Marteau, T. M., Ogilvie, D., Roland, M., Suhrcke, M., & Kelly, M. P. (2011). Judging nudging: can nudging improve population health? BMJ, 342.

3 Mažar, N., & Soman, D. (Eds.). (2022). Behavioral Science in the Wild. University of Toronto Press.

4 Schmidt, R., & Stenger, K. (2021). Behavioral brittleness: The case for strategic behavioral public policy. Behavioural Public Policy, 1-26. Lambe, F., Ran, Y., Jürisoo, M., Holmlid, S., Muhoza, C., Johnson, O., & Osborne, M. (2020). Embracing complexity: A transdisciplinary conceptual framework for understanding behavior change in the context of development-focused interventions. World Development, 126, 104703. Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. Social Science & Medicine, 210, 2-21

5 Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. Annual Review of Psychology, 69(1), 487-510. Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. Psychological Bulletin, 144(12)

6 Muthukrishna, M., & Henrich, J. (2019). A problem in theory. Nature Human Behaviour, 3(3), 221-229. Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. Nature Human Behaviour, 5(8), 980-989. Sanbonmatsu, D. M., & Johnston, W. A. (2019). Redefining science: The impact of complexity on theory development in social and behavioral research. Perspectives on Psychological Science, 14(4)

7 IJzerman, H., Lewis, N.A., Przybylski, A.K. et al. Use caution when applying behavioural science to policy. Nature Human Behaviour 4, 1092–1094 (2020). https://doi.org/10.1038/s41562-020-00990-w

8 Grüne-Yanoff, T. (2012). Old wine in new casks: libertarian paternalism still violates liberal principles. Social Choice and Welfare, 38(4), 635-645. Rizzo, M. J., & Whitman, G. (2020). Escaping paternalism: Rationality, behavioral economics, and public policy. Cambridge University Press.

9 Ewert, B. (2020). Moving beyond the obsession with nudging individual behaviour: Towards a broader understanding of Behavioural Public Policy. Public Policy and Administration, 35(3), 337-360. Lamont, M., Adler, L., Park, B. Y., & Xiang, X. (2017). Bridging cultural sociology and cognitive psychology in three contemporary research programmes. Nature Human Behaviour, 1(12), 866-872. Leggett, W. (2014). The Politics of Behaviour Change: Nudge, Neoliberalism and the State. Policy & Politics 42(1), 3–19.

10 Lorenz-Spreen, P., Geers, M., Pachur, T., Hertwig, R., Lewandowsky, S., & Herzog, S. M. (2021). Boosting people's ability to detect microtargeted advertising. Scientific Reports, 11(1), 1-9. Mills, S. (2022). Personalized nudging. Behavioural Public Policy, 6(1), 150-159. Mohlmann, M. (2021) Algorithmic Nudges Don't Have to Be Unethical. Harvard Business Review. Mills, S. (2022). Personalized Nudging. Behavioural Public Policy, 6(1), 150-159.

11 Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? Behavioral and Brain Sciences, 33(2-3), 61-83.

12 Lewis Jr, N. A. (2021). What counts as good science? How the battle for methodological legitimacy affects public psychology. American Psychologist, 76(8), 1323. Dupree, C. H., & Kraus, M. W. (2022). Psychological science is not race neutral. Perspectives on Psychological Science, 17(1), 270-275.

13 Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? Journal of Economic Perspectives, 21(2), 153-174.

14 Hansen, P. G. (2019). Tools and ethics for applied behavioural insights: The BASIC toolkit. Organisation for Economic Cooperation and Development, OECD.

15 Schmidt, R., & Stenger, K. (2021). Behavioral brittleness: the case for strategic behavioral public policy. Behavioural Public Policy, 1-26.

16 Lepenies, R., Mackay, K., & Quigley, M. (2018). Three challenges for behavioural science and policy: the empirical, the normative and the political. Behavioural Public Policy, 2(2), 174-182.

17 Schmidt, R., & Stenger, K. (2021).

18 Van Rooij, B., & Fine, A. (2021). The behavioral code: The hidden ways the law makes us better or worse. Beacon Press. Andre, P., Haaland, I., Roth, C., & Wohlfart, J. (2021). Narratives about the Macroeconomy (No. 18/21). CEBI Working Paper Series.

19 Dolan, P., Hallsworth, M., Halpern, D., King, D., & Vlaev, I. (2010). MINDSPACE: influencing behaviour for public policy. London: Institute for Government and Cabinet Office.

20 Meder, B., Fleischhut, N., & Osman, M. (2018). Beyond the confines of choice architecture: A critical analysis. Journal of Economic Psychology, 68, 36-44.

21 Thibodeau, P. H., & Boroditsky, L. (2011). Metaphors we think with: The role of metaphor in reasoning. PloS One, 6(2), e16782.

22 Feitsma, J. (2019). Brokering behaviour change: The work of behavioural insights experts in government. Policy & Politics, 47(1), 37-56.

23 Battaglio Jr, R. P., Belardinelli, P., Bellé, N., & Cantarelli, P. (2019). Behavioral public administration ad fontes: A synthesis of research on bounded rationality, cognitive biases, and nudging in public organizations. Public Administration Review, 79(3), 304-320.

24 Schmidt, R. (2022). A model for choice infrastructure: looking beyond choice architecture in Behavioral Public Policy. Behavioural Public Policy, 1-26

25 Soman, D., & Yeung, C. (Eds.). (2020). The behaviourally informed organization. University of Toronto Press.

26 HM Treasury (2020). Magenta Book 2020: Supplementary Guide: Handling Complexity in Policy Evaluation.

27 Boulton, J. G., Allen, P. M., & Bowman, C. (2015). Embracing complexity: Strategic perspectives for an age of turbulence. Oxford University Press.

28 Angeli, F., Camporesi, S., & Dal Fabbro, G. (2021). The COVID-19 wicked problem in public health ethics: conflicting evidence, or incommensurable values? Humanities and Social Sciences Communications, 8(1), 1-8

29 Bak-Coleman, J. B., Alfano, M., Barfuss, W., Bergstrom, C. T., Centeno, M. A., Couzin, I. D., ... & Weber, E. U. (2021). Stewardship of global collective behavior. Proceedings of the National Academy of Sciences, 118(27), e2025764118.

30 Scott, James C. (1998) Seeing like a State. Yale University Press

31 Chater, N., & Loewenstein, G. (2023).

32 Boulton, J. G., Allen, P. M., & Bowman, C. (2015). Embracing complexity: Strategic perspectives for an age of turbulence. OUP Oxford.

33 Schill, C., et al. (2019). A more dynamic understanding of human behaviour for the Anthropocene. Nature Sustainability, 2(12), 1075-1082.

34 DiMaggio, P., & Markus, H. R. (2010). Culture and social psychology: Converging perspectives. Social Psychology Quarterly, 73(4), 347-352.

35 Hallsworth, M. (2017). Rethinking public health using behavioural science. Nature Human Behaviour, 1(9), 612-612.

36 Asano, Y. M., Kolb, J. J., Heitzig, J., & Farmer, J. D. (2021). Emergent inequality and business cycles in a simple behavioral macroeconomic model. Proceedings of the National Academy of Sciences, 118(27).

37 Bak-Coleman, et al. (2021). Stewardship of global collective behavior. Proceedings of the National Academy of Sciences, 118(27), e2025764118.

38 Jones-Rooy, A., & Page, S. E. (2012). The complexity of system effects. Critical Review, 24(3), 313-342.

39 Hawe, P., Shiell, A., & Riley, T. (2009). Theorising interventions as events in systems. American Journal of Community Psychology, 43(3-4), 267-276.

40 Hallsworth, M. (2011) System Stewardship: The future of policymaking? Institute for Government.

41 Rates, C. A., Mulvey, B. K., Chiu, J. L., & Stenger, K. (2022). Examining ontological and self-monitoring scaffolding to improve complex systems thinking with a participatory simulation. Instructional Science, 1-23.

42 Fernandes, L., Morgado, L., Paredes, H., Coelho, A., & Richter, J. (2019). Immersive learning experiences for understanding complex systems. In: iLRN 2019 London-Workshop, Long and Short Paper, Poster, Demos, and SSRiP Proceedings from the Fifth Immersive Learning Research Network Conference (pp. 107-113). Verlag der Technischen Universität Graz.

43 https://www.3ieimpact.org/sites/default/files/2021-07/complexity-blg-Annex1-Checklist_assessing_level_complexity.pdf

44 Treasury, H. M. (2020). Magenta Book Annex: Handling complexity in policy evaluation. Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. Social Science & Medicine, 210, 2-21.

45 Bonell, C., Jamal, F., Melendez-Torres, G. J., & Cummins, S. (2015). 'Dark logic': theorising the harmful consequences of public health interventions. J Epidemiol Community Health, 69(1), 95-98.

46 Kim, D. A., Hwong, A. R., Stafford, D., Hughes, D. A., O'Malley, A. J., Fowler, J. H., & Christakis, N. A. (2015). Social network targeting to maximise population behaviour change: A cluster randomised controlled trial. The Lancet, 386(9989), 145-153.

47 Berry, D. A. (2006). Bayesian clinical trials. Nature Reviews Drug Discovery, 5(1), 27-36.

48 https://www.bi.team/blogs/running-rcts-with-complex-interventions/

49 Volpp, K. G., Terwiesch, C., Troxel, A. B., Mehta, S., & Asch, D. A. (2013, June). Making the RCT more useful for innovation with evidence-based evolutionary testing. In Healthcare (Vol. 1, No. 1-2, pp. 4-7). Elsevier.

50 Kidwell, K. M., & Hyde, L. W. (2016). Adaptive interventions and SMART designs: application to child behavior research in a community setting. American Journal of Evaluation, 37(3), 344-363.

51 Caria, S., Kasy, M., Quinn, S., Shami, S., & Teytelboym, A. (2020). An adaptive targeted field experiment: Job search assistance for refugees in Jordan. SSRN.

52 https://www.cecan.ac.uk/wp-content/uploads/2020/08/EPPN-No-03-Agent-Based-Modelling-for-Evaluation.pdf

53 Schlüter, M., et al. (2017). A framework for mapping and comparing behavioural theories in models of social-ecological systems. Ecological Economics, 131, 21-35. Wijermans, N., Boonstra, W. J., Orach, K., Hentati-Sundberg, J., & Schlüter, M. (2020). Behavioural diversity in fishing—Towards a next generation of fishery models. Fish and Fisheries, 21(5), 872-890.

54 Schill, C., et al. (2019). A more dynamic understanding of human behaviour for the Anthropocene. Nature Sustainability, 2(12), 1075-1082.

55 Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological Science, 22(11), 1359-1366; Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. Annual Review of Psychology, 69, 511-534.

56 Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. Psychological Bulletin, 144(12), 1325.

57 Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. Proceedings of the National Academy of Sciences, 113(23), 6454-6459. Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. Nature Human Behaviour, 5(8), 980-989.

58 McShane, B. B., Tackett, J. L., Böckenholt, U., & Gelman, A. (2019). Large-scale replication projects in contemporary psychological research. The American Statistician, 73(sup1), 99-105.

59 Sanbonmatsu, D. M., Cooley, E. H., & Butner, J. E. (2021). The Impact of Complexity on Methods and Findings in Psychological Science. Frontiers in Psychology, 4028.

60 Cartwright, N., & Hardie, J. (2012). Evidence-based policy: A practical guide to doing it better. Oxford University Press. Soman, D. & Mazar, N. (2022) The Science of Translation and Scaling. In: Mazar and Soman (eds.) Behavioral Science in the Wild, University of Toronto Press, pp.5-19.

61 https://osf.io/zuh93/

62 Landy, J. F., et al. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. Psychological Bulletin, 146(5), 451.

63 Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. Proceedings of the National Academy of Sciences, 113(23), 6454-6459.

64 Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implementation Science, 4(1), 1-15.

65 Mazar, N. & Soman, D. (eds.) (2022) Behavioral Science in the Wild, University of Toronto Press

66 Brenninkmeijer, J., Derksen, M., Rietzschel, E., Vazire, S., & Nuijten, M. (2019). Informal laboratory practices in psychology. Collabra: Psychology, 5(1).

67 McShane, B. B., Tackett, J. L., Böckenholt, U., & Gelman, A. (2019). Large-scale replication projects in contemporary psychological research. The American Statistician, 73(sup1), 99-105.

68 Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. Psychonomic bulletin & review, 26(5), 1596-1618. Borsboom, D., van der Maas, H. L., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. Perspectives on Psychological Science, 16(4), 756-766.

69 Sanbonmatsu, D. M., Cooley, E. H., & Butner, J. E. (2021). The Impact of Complexity on Methods and Findings in Psychological Science. Frontiers in Psychology, 4028.

70 Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. Psychonomic Bulletin & Review, 26(5), 1596-1618.

71 Fried, E. I. (2020). Theories and models: What they are, what they are for, and what they are about. Psychological Inquiry, 31(4), 336-344.

72 Muthukrishna, M., & Henrich, J. (2019). A problem in theory. Nature Human Behaviour, 3(3), 221-229.

73 http://behavioralscientist.org/there-is-more-to-behavioral-science-than-biases-and-fallacies/

74 Muthukrishna, M., & Henrich, J. (2019). A problem in theory. Nature Human Behaviour, 3(3), 221-229.

75 Abner, G. B., Kim, S. Y., & Perry, J. L. (2017). Building evidence for public human resource management: Using middle range theory to link theory and data. Review of Public Personnel Administration, 37(2), 139-159.

76 Moore, L. F., Johns, G., & Pinder, C. C. (1980). Toward middle range theory. Middle range theory and the study of organizations, 1-16.

77 Sanbonmatsu, D. M., Cooley, E. H., & Butner, J. E. (2021). The Impact of Complexity on Methods and Findings in Psychological Science. Frontiers in Psychology, 4028.

78 Berkman, E. T., & Wilson, S. M. (2021). So useful as a good theory? The practicality crisis in (social) psychological theory. Perspectives on Psychological Science, 16(4), 864-874.

79 Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. Behavioral and Brain Sciences, 43.

80 Callaway, F., Hardy, M., & Griffiths, T. (2022). Optimal nudging for cognitively bounded agents: A framework for modeling, predicting, and controlling the effects of choice architectures. Working Paper.

81 Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. Perspectives on Psychological Science, 7(5), 411-426.

82 Henriksen, K., & Kaplan, H. (2003). Hindsight bias, outcome knowledge and adaptive learning. BMJ Quality & Safety, 12(suppl 2), ii46-ii50. Bukszar, E., & Connolly, T. (1988). Hindsight bias and strategic choice: Some problems in learning from experience. Academy of Management Journal, 31(3), 628-641. DellaVigna, S., Pope, D., & Vivalt, E. (2019). Predict science to improve science. Science, 366(6464), 428-429.

83 DellaVigna, S., Pope, D., & Vivalt, E. (2019). Predict science to improve science. Science, 366(6464), 428-429.

84 Munnich, E., & Ranney, M. A. (2019). Learning from surprise: Harnessing a metacognitive surprise signal to build and adapt belief networks. Topics in Cognitive Science, 11(1), 164-177.

85 Dimant, E., Clemente, E. G., Pieper, D., Dreber, A., & Gelfand, M. (2022). Politicizing mask-wearing: predicting the success of behavioral interventions among republicans and democrats in the US. Scientific Reports, 12(1), 1-12. DellaVigna, S., & Linos, E. (2022). RCTs to scale: Comprehensive evidence from two nudge units. Econometrica, 90(1), 81-116.

86 Ackerman, R., Bernstein, D. M., & Kumar, R. (2020). Metacognitive hindsight bias. Memory & Cognition, 48(5), 731-744.

87 Pezzo, M. (2003). Surprise, defence, or making sense: What removes hindsight bias? Memory, 11(4-5), 421-441.

88 Schmidt, R., & Stenger, K. (2021). Behavioral brittleness: the case for strategic behavioral public policy. Behavioural Public Policy, 1-26.

89 Dorison, C. A., & Heller, B. H. (2022). Observers penalize decision makers whose risk preferences are unaffected by loss–gain framing. Journal of Experimental Psychology: General, 151(9).

90 Porter, T., Elnakouri, A., Meyers, E. A., Shibayama, T., Jayawickreme, E., & Grossmann, I. (2022). Predictors and consequences of intellectual humility. Nature Reviews Psychology, 1(9), 524-536.

91 Hallsworth, M., Egan, M., Rutter, J., & McCrae, J. (2018). Behavioural government: Using behavioural science to improve how governments make decisions. London: Institute for Government.

92 Van Bavel, R., & Dessart, F. J. (2018). The case for qualitative methods in behavioural studies for EU policy-making. Publications Office of the European Union: Luxembourg. https://www.povertyactionlab.org/blog/8-11-21/strengthening-randomized-evaluations-through-incorporating-qualitative-research-part-1

93 Walton, G. M., & Wilson, T. D. (2018). Wise interventions: Psychological remedies for social and personal problems. Psychological Review, 125(5), 617.

94 Lewis Jr, N. A. (2021). What counts as good science? How the battle for methodological legitimacy affects public psychology. American Psychologist, 76.

95 Lamont, M., Adler, L., Park, B. Y., & Xiang, X. (2017). Bridging cultural sociology and cognitive psychology in three contemporary research programmes. Nature Human Behaviour, 1(12), 866-872. Vaisey, S. (2009). Motivation and justification: A dual-process model of culture in action. American journal of sociology, 114(6), 1675-1715.

96 Richardson, L., & John, P. (2021). Co-designing behavioural public policy: lessons from the field about how to 'nudge plus'. Evidence & Policy, 17(3), 405-422. Hallsworth, M. & Kirkman, E. (2020).

97 Banerjee, S., & John, P. (2020). Nudge plus: incorporating reflection into behavioral public policy. Behavioural Public Policy, 1-16. Reijula, S., & Hertwig, R. (2022). Self-nudging and the citizen choice architect. Behavioural Public Policy, 6(1), 119-149. Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. Perspectives on Psychological Science, 12(6), 973-986.

98 Hertwig, R. (2017). When to consider boosting: some rules for policy-makers. Behavioural Public Policy, 1(2), 143-161. See also: Grüne-Yanoff, T., Marchionni, C., & Feufel, M. A. (2018). Toward a framework for selecting behavioural policies: How to choose between boosts and nudges. Economics & Philosophy, 34(2), 243-266

99 Sims, A., & Müller, T. M. (2019). Nudge versus boost: A distinction without a normative difference. Economics & Philosophy, 35(2), 195-222.

100 Big-data studies of human behaviour need a common language. Nature July 8, 2021. Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. Perspectives on Psychological Science, 12(6), 1100-1122.

101 Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 113(523), 1228-1242. Künzel, S. R., Sekhon, J. S., Bickel, P. J., & Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. Proceedings of the National Academy of Sciences, 116(10), 4156-4165.

102 Todd-Blick, A., Spurlock, C. A., Jin, L., Cappers, P., Borgeson, S., Fredman, D., & Zuboy, J. (2020). Winners are not keepers: Characterizing household engagement, gains, and energy patterns in demand response using machine learning in the United States. Energy Research & Social Science, 70, 101595.

103 Mills, S. (2022). Personalized nudging. Behavioural Public Policy, 6(1), 150-159.

104 Soman, D., & Hossain, T. (2021). Successfully scaled solutions need not be homogenous. Behavioral Public Policy, 5(1), 80–9.

105 Mills, S. (2022). Personalized nudging. Behavioural Public Policy, 6(1), 150-159. Lorenz-Spreen, P., Geers, M., Pachur, T., Hertwig, R., Lewandowsky, S., & Herzog, S. M. (2021). Boosting people's ability to detect microtargeted advertising. Scientific Reports, 11(1), 1-9.

106 Kozyreva, A., Lorenz-Spreen, P., Hertwig, R., Lewandowsky, S., & Herzog, S. M. (2021). Public attitudes towards algorithmic personalization and use of personal data online: Evidence from Germany, Great Britain, and the United States. Humanities and Social Sciences Communications, 8(1), 1-11.

107 De Jonge, P., Verlegh, P. & Zeelenberg, M. (2022) If You Want People to Accept Your Intervention, Don't Be Creepy. In: Soman, D. & Mazar, N. (eds) Behavioral Science in the Wild, University of Toronto Press, pp.284-291.

108 https://thedecisionlab.com/insights/technology/this-is-personal-the-dos-and-donts-of-personalization-in-tech

109 Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. Proceedings of the National Academy of Sciences, 114(48), 12714-12719.

110 Lorenz-Spreen, P., Geers, M., Pachur, T., Hertwig, R., Lewandowsky, S., & Herzog, S. M. (2021). Boosting people's ability to detect microtargeted advertising. Scientific Reports, 11(1), 1-9.

111 Lewis Jr, N. A. (2021). What counts as good science? How the battle for methodological legitimacy affects public psychology. American Psychologist, 76(8), 1323.

112 Nagel, T. (1986) The View from Nowhere. Sugden, R. (2013). The behavioural economist and the social planner: to whom should behavioural welfare economics be addressed?. Inquiry, 56(5), 519-538.

113 Liscow, Z., & Markovits, D. (2022). Democratizing Behavioral Economics. Yale Journal on Regulation, 39(1217).

114 Roberts, S. O., Bareket-Shavit, C., Dollins, F. A., Goldie, P. D., & Mortenson, E. (2020). Racial inequality in psychological research: Trends of the past and recommendations for the future. Perspectives on Psychological Science, 15(6), 1295-1309.

115 https://gocommonthread.com/work/global-gavi/bi

116 Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. Behavioral and brain sciences, 33(2-3), 61-83. Cheon, B. K., Melani, I., & Hong, Y. Y. (2020). How USA-centric is psychology? An archival study of implicit assumptions of generalizability of findings to human nature based on origins of study samples. Social Psychological and Personality Science, 11(7), 928-937.

117 Dupree, C. H., & Kraus, M. W. (2022). Psychological science is not race neutral. Perspectives on Psychological Science, 17(1), 270-275.

118 Sendhil Mullainathan, keynote address to the Society of Judgment and Decision Making Annual Conference, 2022.

119 https://chicagobeyond.org/researchequity/

120 https://www.poverty-action.org/blog/locally-grounded-research-strengthening-partnerships-advance-science-and-impact-development

121 https://www.bi.team/blogs/increasing-economic-mobility-in-us-cities/. OECD (2017) Behavioural Insights and Public Policy: Lessons from Around the World; https://www.un.org/en/content/behaviouralscience/

122 Hallsworth, M., Snijders, V., Burd, H., Prestt, J., Judah, G., Huf, S., & Halpern, D. (2016). Applying behavioral insights: Simple ways to improve health outcomes. World Innovation Summit for Health.

123 Damgaard, M. T., & Nielsen, H. S. (2018). Nudging in education. Economics of Education Review, 64, 313-342.

124 Ferrari, L., Cavaliere, A., De Marchi, E., & Banterle, A. (2019). Can nudging improve the environmental impact of food supply chain? A systematic review. Trends in Food Science & Technology, 91, 184-192. Byerly, H., et al. (2018). Nudging pro-environmental behavior: evidence and opportunities. Frontiers in Ecology and the Environment, 16(3), 159-168.

125 Dyson, P. & Sutherland, R. (2021) Transport for humans: Are we nearly there yet? London publishing.

126 Cadario, R., & Chandon, P. (2020). Which healthy eating nudges work best? A meta-analysis of field experiments. Marketing Science, 39(3), 465-486.

127 Money Advice Service, Behavioural Insights Team, and Ipsos MORI (2018) A behavioural approach to managing money: Ideas and results from the Financial Capability Lab.

128 Benartzi, S., et al. (2017). Should governments invest more in nudging? Psychological Science, 28(8), 1041-1055.

129 Chater, N., & Loewenstein, G. (2023).

130 Sanders, M., Snijders, V., & Hallsworth, M. (2018). Behavioural science and policy: Where are we now and where are we going?. Behavioural Public Policy, 2(2), 144-167.

131 Sanders, M., Snijders, V., & Hallsworth, M. (2018).

132 See, for example, Hansen, P. G. (2019). Tools and ethics for applied behavioural insights: The BASIC toolkit. Organisation for Economic Cooperation and Development, OECD; Datta, S., & Mullainathan, S. (2014). Behavioral design: a new approach to development policy. Review of Income and Wealth, 60(1), 7-35.

133 Schmidt, R., & Stenger, K. (2021). Behavioral brittleness: the case for strategic behavioral public policy. Behavioural Public Policy, 1-26.

134 Hallsworth, M., Berry, D., Sanders, M., Sallis, A., King, D., Vlaev, I., & Darzi, A. (2015). Stating appointment costs in SMS reminders reduces missed hospital appointments: findings from two randomised controlled trials. PloS one, 10(9), e0137306.

135 https://www.nsw.gov.au/behavioural-insights-unit/blog/reducing-missed-hospital-appointments-better-text-messages; Berliner Senderey, A., Kornitzer, T., Lawrence, G., Zysman, H., Hallak, Y., Ariely, D., & Balicer, R. (2020). It's how you say it: Systematic A/B testing of digital messaging cut hospital no-show rates. PloS one, 15(6), e0234817.

136 Mills, S. (2020). Nudge/sludge symmetry: on the relationship between nudge and sludge and the resulting ontological, normative and transparency implications. Behavioural Public Policy, 1-24.

137 Schmidt, R., & Stenger, K. (2021). Behavioral brittleness: The case for strategic behavioral public policy. Behavioural Public Policy, 1-26.

138 https://medium.com/@DavePerrott/is-applied-behavioural-science-reaching-a-local-maximum-538b536f7e7d

139 Hansen, P. G. (2018). What are we forgetting? Behavioural Public Policy, 2(2), 190-197.

140 Costa, E., King, K., Dutta, R., & Algate, F. (2016). Applying behavioural insights to regulated markets. The Behavioural Insights Team.

141 https://www.bi.team/blogs/behaviour-change-and-the-new-sugar-tax/

142 The Behavioural Insights Team (2020) The Behavioural Economy.

143 Ewert, B., Loer, K., & Thomann, E. (2021). Beyond nudge: Advancing the state-of-the-art of behavioural public policy and administration. Policy & Politics, 49(1), 3-23.

144 Halpern, D. (2015). Inside the nudge unit: How small changes can make a big difference. WH Allen.

145 Hansen, P. G. (2018). What are we forgetting? Behavioural Public Policy, 2(2), 190-197.

146 Soman, D., & Yeung, C. (eds.) (2020). The behaviourally informed organization. University of Toronto Press.

147 Dolan, P., Hallsworth, M., Halpern, D., King, D., & Vlaev, I. (2010). MINDSPACE: influencing behaviour for public policy. London: Institute for Government and Cabinet Office.

148 Meder, B., Fleischhut, N., & Osman, M. (2018). Beyond the confines of choice architecture: a critical analysis. Journal of Economic Psychology, 68, 36-44.

149 Schmidt, R. (2022). A model for choice infrastructure: looking beyond choice architecture in Behavioral Public Policy. Behavioural Public Policy, 1-26. However, the notion of 'infrastructure' has connotations of physical rigidity, which may sit badly with the dynamic nature of human infrastructure (interactions and relationships).

150 Thibodeau, P. H., & Boroditsky, L. (2011). Metaphors we think with: The role of metaphor in reasoning. PloS One, 6(2), e16782.

151 Khan, Z. & Newman, L. (2021). Building behavioral science in an organization. Action Design Press.

152 Wendel, S. (October 5, 2020). Who is doing applied behavioural science? Results from a Global Survey of Behavioural Teams. https://behavioralscientist.org/who-is-doing-applied-behavioral-science-results-from-a-global-survey-of-behavioral-teams/

153 https://gocommonthread.com/work/global-gavi-bi/. Halpern, D. (2015) Inside the Nudge Unit. WH Allen. Angawi, A. & Hasanain, W. (2018) The Nuts and Bolts of Behavioral Insights Units. In: The Behavioral Economics Guide 2018, ed. Alain Sasom. Robertson, T., Darling, M., Leifer, J., Footer, O., & Gordski, D. (2017). Behavioral design teams: The next frontier in clinical delivery innovation. Issue Brief, 2017, 1-16.

154 Battaglio Jr, R. P., Belardinelli, P., Bellé, N., & Cantarelli, P. (2019). Behavioral public administration ad fontes: A synthesis of research on bounded rationality, cognitive biases, and nudging in public organizations. Public Administration Review, 79(3), 304-320.

155 https://hbr.org/2017/10/why-coos-should-think-like-behavioral-economists

156 Soman, D. (2020). The Science of Using Behavioral Science. In: The Behaviourally Informed Organization (pp. 3-22). University of Toronto Press, pp.12-13.

157 Mayer, S., Shah, R., & Kalil, A. (2021). How cognitive biases can undermine program scale-up decisions. In: The Scale-Up Effect in Early Childhood and Public Policy (pp. 41-57). Routledge.

158 CPI (2021). Human Learning Systems Report, p.62

159 Behavioral Insights Team (2018). Behavioral Government.

160 Feng, Kim & Soman (2020) have a related but different diagram, which focuses on what issues the behavioral strategies are applied to. Feng, B., Kim, M., & Soman, D. (2020). Embedding Behavioral Insights in Organizations. In: The Behaviourally Informed Organization (pp. 23-40). University of Toronto Press.

161 To minimize complexity, we are combining 'knowledge', referring to awareness of the concepts of behavioral science, and 'capacity', referring to the ability to implement them - e.g., construct interventions and test them.

162 Obviously, this diagram greatly simplifies matters. It does not fully account for the relationships between organizations (as in those between government departments). The World Bank adds a third category, 'networked', to represent resources that are diffused but coordinated Afif, Z., Islan, W. W., Calvo-Gonzalez, O., & Dalton, A. (2018). Behavioral science around the world: Profiles of 10 countries. World Bank Group: Mind, Behavior, and Development Unit.

163 Feng, B., Kim, M., & Soman, D. (2020). Embedding Behavioral Insights in Organizations. In: The Behaviourally Informed Organization (pp. 23-40). University of Toronto Press, p.31.

164 Behavioral Insights Team (2018) Behavioral Government.

165 Herd, P. & Moynihan, D (2018) Administrative Burden: Policymaking by other means. Russell Sage Foundation. Sunstein, C. R. (2022). Sludge audits. Behavioural public policy, 6(4), 654-673.

166 Cantarelli, P., Bellé, N., & Belardinelli, P. (2020). Behavioral public HR: Experimental evidence on cognitive biases and debiasing interventions. Review of Public Personnel Administration, 40(1), 56-81.

167 Cantarelli, P., Bellé, N., & Belardinelli, P. (2020).

168 Feitsma, J. (2019). Brokering behaviour change: the work of behavioural insights experts in government. Policy & Politics, 47(1), 37-56.

169 Hallsworth, M., & Kirkman, E. (2020). p.192.

170 Soman, D. (2020). The Science of Using Behavioral Science. In: The Behaviourally Informed Organization (pp. 3-22). University of Toronto Press, p.18.

171 Soman, D., & Yeung, C. (2020), p.265.

172 Soman, D., & Yeung, C. (2020), p.257.

173 Feitsma, J. N. P. (2019). Inside the behavioural state. Eleven International Publishing.

174 Soman, D., & Yeung, C. (2020), p.265.

175 Feitsma, J. N. P. (2019). Inside the behavioural state. Eleven International Publishing, p.109.

176 Soman and Yeung (2020) put forward the principle of the 'behaviourally-informed organization'. I support and draw on this work. However, for my purposes the term "informed" introduces an ambiguity. The processes in the 'nudged organization' may be seen as informed by behavioral science, even though few people may be informed of the principles being applied.

177 Ewert, B., & Loer, K. (2021). Advancing behavioural public policies: in pursuit of a more comprehensive concept. Policy & Politics, 49(1), 25-47.

178 Schmidt, R. (2022). A model for choice infrastructure: looking beyond choice architecture in Behavioral Public Policy. Behavioural Public Policy, 1-26

179 Obviously, the focus here will change depending on whether you are looking to build behavioral science knowledge and capacity (i.e., move further right in the diagram). There are similarities here to the distinction that Lourenco et al. (2016) make between 'behaviorally informed' initiatives (designed after an explicit review of evidence on behavior) and 'behaviorally aligned' initiatives (initiatives that do not rely on behavioral evidence but which can be found to be in line with it).

180 Schmidt, R. (2022). A model for choice infrastructure: looking beyond choice architecture in Behavioral Public Policy. Behavioural Public Policy, 1-26

181 Soman, D., & Yeung, C. (eds.) (2020). The behaviourally informed organization. University of Toronto Press.

182 Soman, D., & Yeung, C. (2020), p.284.

183 Grimmelikhuijsen, S., Jilke, S., Olsen, A. L., & Tummers, L. (2017). Behavioral public administration: Combining insights from public administration and psychology. Public Administration Review, 77(1), 45-56. Donohue, K., Özer, Ö., & Zheng, Y. (2020). Behavioral operations: Past, present, and future. Manufacturing & Service Operations Management, 22(1), 191-202.

184 Sunstein, C. R. (2022). Sludge audits. Behavioural public policy, 6(4), 654-673.

185 HM Treasury (2020) Magenta Book 2020: Supplementary Guide: Handling Complexity in Policy Evaluation.

186 Savona, N., Thompson, C., Rutter, H., & Cummins, S. (2017). Exposing complexity as a smokescreen: A qualitative analysis. The Lancet, 390, S3.

187 Hallsworth, M. & Kirkman, E. (2020), pp.185-186.

188 Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a general theory of planning. Policy Sciences, 4(2), 155-169.

189 Bak-Coleman, J. B., et al. (2021). Stewardship of global collective behavior. Proceedings of the National Academy of Sciences, 118(27), e2025764118.

190 Krakauer, D. & West, G. (2021) The Complex Alternative: Complexity Scientists on the COVID-19 Pandemic. The Santa Fe Institute.

191 Angeli, F., Camporesi, S., & Dal Fabbro, G. (2021). The COVID-19 wicked problem in public health ethics: conflicting evidence, or incommensurable values? Humanities and Social Sciences Communications, 8(1), 1-8.

192 Bak-Coleman, J. B., et al. (2021).

193 Endo, A. (2020). Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. Wellcome Open Research, 5.

194 Solé, R., & Elena, S. F. (2018). Viruses as complex adaptive systems (Vol. 15). Princeton University Press. https://stemacademicpress.com/stem-volumes-covid-19

195 Bak-Coleman, J. B., et al. (2021).

196 Turcotte-Tremblay, A. M., Gali, I. A. G., & Ridde, V. (2021). The unintended consequences of COVID-19 mitigation measures matter: practical guidance for investigating them. BMC Medical Research Methodology, 21(1), 1-17.

197 Lambe, F., Ran, Y., Jürisoo, M., Holmlid, S., Muhoza, C., Johnson, O., & Osborne, M. (2020). Embracing complexity: A transdisciplinary conceptual framework for understanding behavior change in the context of development-focused interventions. World Development, 126, 104703.

198 Ruth Schmidt and Kateleyn Stenger sum up the challenges in terms of 'contextual brittleness', 'systemic brittleness', and 'anticipatory brittleness' in an excellent article. Schmidt, R., & Stenger, K. (2021). Behavioral brittleness: the case for strategic behavioral public policy. Behavioural Public Policy, 1-26.

199 Government Communication Service (2021). The Principles of Behaviour Change Communications, p.34 At: https://gcs.civilservice.gov.uk/publications/the-principles-of-behaviour-change-communications/

200 Center for Public Innovation (2020) Human Learning Systems Report, p.76.

201 Schmidt, R., & Stenger, K. (2021). Behavioral brittleness: the case for strategic behavioral public policy. Behavioural Public Policy, 1-26. Altmann, S., Grunewald, A., & Radbruch, J. (2021). Interventions and Cognitive Spillovers, The Review of Economic Studies, rdab087, https://doi.org/10.1093/restud/rdab087

202 Fisman, R. & Golden, M. (2017). How to fight corruption. Science, 356(6340): 803–804.

203 Taylor, R. L. (2019). Bag leakage: The effect of disposable carryout bag regulations on unregulated bags. Journal of Environmental Economics and Management, 93, 254-271.

204 Medina, P. C. (2021). Side effects of nudging: Evidence from a randomized intervention in the credit card market. The Review of Financial Studies, 34(5), 2580-2607.

205 Cook, R., & Rasmussen, J. (2005). "Going solid": a model of system dynamics and consequences for patient safety. BMJ Quality & Safety, 14(2), 130-134.

206 Bak-Coleman, J. B., et al. (2021).

207 Boulton, J. G., Allen, P. M., & Bowman, C. (2015). Embracing complexity: Strategic perspectives for an age of turbulence. OUP Oxford.

208 Jones-Rooy, A., & Page, S. E. (2012). The complexity of system effects. Critical Review, 24(3), 313-342.

209 Dentoni, D., Bitzer, V., & Schouten, G. (2018). Harnessing wicked problems in multi-stakeholder partnerships. Journal of Business Ethics, 150(2), 333-356.

210 Barak-Corren, N., & Kariv-Teitelbaum, Y. (2021). Behavioral responsive regulation: Bringing together responsive regulation and behavioral public policy. Regulation & Governance, 15, S163-S182.

211 Hallsworth, M. (2012). How complexity economics can improve government: rethinking policy actors, institutions and structures. Complex new world: translating new economic thinking into public policy, London: IPPR (Institute for Public Policy Research), 39-49.

212 Gras, D., Conger, M., Jenkins, A., & Gras, M. (2020). Wicked problems, reductive tendency, and the formation of (non-) opportunity beliefs. Journal of Business Venturing, 35(3).

213 Gras, D., Conger, M., Jenkins, A., & Gras, M. (2020).

214 Dunlop, C. A., & Radaelli, C. M. (2015). Overcoming Illusions of Control: How to Nudge and Teach

Regulatory Humility. In A. Alemanno & A. L. Sibony (eds.) Nudge and the Law. A European Perspective (pp. 139-160). Oxford/London: Hart.

215 Scott, James C. (1998) Seeing like a State. Yale University Press

216 Boulton, J. G., Allen, P. M., & Bowman, C. (2015). Embracing complexity: Strategic perspectives for an age of turbulence. OUP Oxford.

217 I am aware of the argument that many social behaviors are 'complex contagions' and therefore require 'contact with multiple sources of reinforcement in order to be transmitted', as opposed to viruses, which are 'simple contagions' that may only require a single contact from a single source. Centola, D. (2018). How Behavior Spreads, p.37.

218 Macy, M., Deri, S., Ruch, A., & Tong, N. (2019). Opinion cascades and the unpredictability of partisan polarization. Science Advances, 5(8), eaax0754.

219 Schill, C., et al. (2019). A more dynamic understanding of human behaviour for the Anthropocene. Nature Sustainability, 2(12), 1075-1082.; DiMaggio, P., & Markus, H. R. (2010). Culture and social psychology: Converging perspectives. Social Psychology Quarterly, 73(4), 347-352.

220 Chater, N., & Loewenstein, G. (2023)

221 Schill, C., et al. (2019).

222 There are similarities here to the idea of "Coleman's boat" in sociology. Ramström, G. (2018). Coleman's boat revisited: Causal sequences and the micro-macro link. Sociological Theory, 36(4), 368-391.

223 Abson, D. J., et al. (2017). Leverage points for sustainability transformation. Ambio, 46(1), 30-39.

224 Currin, C. B., Vera, S. V., & Khaledi-Nasab, A. (2022). Depolarization of echo chambers by random dynamical nudge. Scientific Reports, 12(1), 1-13.

225 Andreoni, J., Nikiforakis, N., & Siegenthaler, S. (2021). Predicting social tipping and norm change in controlled experiments. Proceedings of the National Academy of Sciences, 118(16), e2014893118.

226 Daniels, B. C., Krakauer, D. C., & Flack, J. C. (2017). Control of finite critical behaviour in a small-scale social system. Nature Communications, 8(1), 1-8.

227 Alexander, M., Forastiere, L., Gupta, S., & Christakis, N. A. (2022). Algorithms for seeding social networks can enhance the adoption of a public health intervention in urban India. Proceedings of the National Academy of Sciences, 119(30), e2120742119.

228 Hallsworth, M. (2017). Rethinking public health using behavioural science. Nature Human Behaviour, 1(9), 612-612.

229 https://ondrugs.substack.com/p/president-bidens-scheduling-directive

230 https://www.bbc.com/news/health-40444460

231 Jones-Rooy, A., & Page, S. E. (2012). The complexity of system effects. Critical Review, 24(3), 313-342.

232 Holland, J. H. (1996). Hidden order: How adaptation builds complexity. Addison Wesley Longman Publishing Co., Inc.; Holland, J. H. (2000). Emergence: From chaos to order. OUP Oxford.

233 Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. Annual Review of Psychology, 62, 451-482.

234 Akerlof, G. A., & Shiller, R. J. (2010). Animal spirits: How human psychology drives the economy, and why it matters for global capitalism. Princeton University Press.

235 Asano, Y. M., Kolb, J. J., Heitzig, J., & Farmer, J. D. (2021). Emergent inequality and business cycles in a simple behavioral macroeconomic model. Proceedings of the National Academy of Sciences, 118(27).

236 Jones-Rooy, A., & Page, S. E. (2012). The complexity of system effects. Critical Review, 24(3), 313-342.

237 Bak-Coleman, J. B., et al. (2021).

238 Hallsworth, M. & Kirkman, E. (2020) Behavioral Insights. MIT Press.

239 https://tfl.gov.uk/info-for/media/press-releases/2022/february/new-tfl-data-shows-success-of-innovative-pedestrian-priority-traffic-signals

240 Hawe, P., Shiell, A., & Riley, T. (2009). Theorising interventions as events in systems. American Journal of Community Psychology, 43(3-4), 267-276.

241 Doctor, J. N., Wakker, P. P., & Wang, T. V. (2020). Economists' views on the ergodicity problem. Nature Physics, 16(12), 1168-1168; Thomas, D. H., & Jona, L. (2018). 'Good Nudge Lullaby': Choice Architecture and Default Bias Reinforcement. The Economic Journal, 128(610), 1180-1206.

242 Hallsworth, M. (2011) System Stewardship. Institute for Government.

243 HM Treasury (2020). The Magenta Book: Supplementary Guide: Handling Complexity in Policy Evaluation.

244 Rates, C. A., Mulvey, B. K., Chiu, J. L., & Stenger, K. (2022). Examining ontological and self-monitoring scaffolding to improve complex systems thinking with a participatory simulation. Instructional Science, 1-23.

245 Fernandes, L., Morgado, L., Paredes, H., Coelho, A., & Richter, J. (2019). Immersive learning experiences for understanding complex systems. In iLRN 2019 London-Workshop, Long and Short Paper, Poster, Demos, and SSRiP Proceedings from the Fifth Immersive Learning Research Network Conference (pp. 107-113). Verlag der Technischen Universität Graz.

246 https://www.3ieimpact.org/sites/default/files/2021-07/complexity-blg-Annex1-Checklist_assessing_level_complexity.pdf

247 Hallsworth, M. & Kirkman, E. (2020). Behavioral Insights. MIT Press.

248 HM Treasury (2020) The Magenta Book: Supplementary Guide: Handling Complexity in Policy Evaluation.

249 Munafò, M. R., et al. (2017). A manifesto for reproducible science. Nature Human Behaviour, 1(1), 1-9.

250 Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. Social Science & Medicine, 210, 2-21.

251 HM Treasury (2020) The Magenta Book: Supplementary Guide: Handling Complexity in Policy Evaluation.

252 Boulton, J. G., Allen, P. M., & Bowman, C. (2015). Embracing complexity: Strategic perspectives for an age of turbulence. OUP Oxford, p.189.

253 Robinson, C. D., Chande, R., Burgess, S., & Rogers, T. (2021). Parent Engagement Interventions are Not Costless: Opportunity Cost and Crowd Out of Parental Investment. Educational Evaluation and Policy Analysis DOI: 10.3102/01623737211030492

254 Note that this is a slightly different challenge from understanding why nudges backfire or fail in general, since that can happen in a relatively simple situation that someone has just misinterpreted. Osman, M., McLachlan, S., Fenton, N., Neil, M., Löfstedt, R., & Meder, B. (2020). Learning from behavioural changes that fail. Trends in Cognitive Sciences. December 2020, Vol. 24, No. 12

255 Bonell, C., Jamal, F., Melendez-Torres, G. J., & Cummins, S. (2015). 'Dark logic': Theorising the harmful consequences of public health interventions. J Epidemiol Community Health, 69(1), 95-98.

256 Catlow, J., Bhardwaj-Gosling, R., Sharp, L., Rutter, M. D., & Sniehotta, F. F. (2022). Using a dark logic model to explore adverse effects in audit and feedback: A qualitative study of gaming in colonoscopy. BMJ Quality & Safety, 31(10), 704-715.

257 Schmidt, R., & Stenger, K. (2021). Behavioral brittleness: the case for strategic behavioral public policy. Behavioural Public Policy, 1-26.

258 Johnson, A. L., et al. (2020). Increasing the impact of randomized controlled trials: an example of a hybrid effectiveness–implementation design in psychotherapy research. Translational Behavioral Medicine, 10(3), 629-636.

259 Centola, D. (2018) How Behavior Spreads: The Science of Complex Contagions. Princeton University Press. p.63.

260 Kim, D. A., Hwong, A. R., Stafford, D., Hughes, D. A., O'Malley, A. J., Fowler, J. H., & Christakis, N. A. (2015). Social network targeting to maximise population behaviour change: a cluster randomised controlled trial. The Lancet, 386(9989), 145-153.

261 Beaman, L., BenYishay, A., Magruder, J., & Mobarak, A. M. (2021). Can network theory-based targeting increase technology adoption? American Economic Review, 111(6), 1918-43.

262 Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2019). Using gossips to spread information: Theory and evidence from two randomized controlled trials. The Review of Economic Studies, 86(6), 2453-2490.

263 Education Endowment Fund (2017) Evaluation of Complex Whole-School Interventions: Methodological and Practical Considerations.

264 Berry, D. A. (2006). Bayesian clinical trials. Nature Reviews Drug Discovery, 5(1), 27-36.

265 Boulton, J. G., Allen, P. M., & Bowman, C. (2015). Embracing complexity: Strategic perspectives for an age of turbulence. OUP Oxford, p.189. HM Treasury (2020) The Magenta Book: Supplementary Guide: Handling Complexity in Policy Evaluation.

266 https://www.bi.team/blogs/running-rcts-with-complex-interventions/

267 Collins, L. M., Murphy, S. A., & Strecher, V. (2007). The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. American Journal of Preventive Medicine, 32(5), S112-S118.

268 Marinelli, H. A., Berlinski, S., & Busso, M. (2021). Remedial education: Evidence from a sequence of experiments in Colombia. Journal of Human Resources, 0320-10801R2.

269 Volpp, K. G., Terwiesch, C., Troxel, A. B., Mehta, S., & Asch, D. A. (2013, June). Making the RCT more useful for innovation with evidence-based evolutionary testing. Healthcare 1(1-2), 4-7.

270 Kidwell, K. M., & Hyde, L. W. (2016). Adaptive interventions and SMART designs: application to child behavior research in a community setting. American Journal of Evaluation, 37(3), 344-363. Klasnja, P., Hekler, E. B., Shiffman, S., Boruvka, A., Almirall, D., Tewari, A., & Murphy, S. A. (2015). Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. Health Psychology, 34(S), 1220.

271 Kasy, M., & Sautmann, A. (2021). Adaptive treatment assignment in experiments for policy choice. Econometrica, 89(1), 113-132.

272 Caria, S., Kasy, M., Quinn, S., Shami, S., & Teytelboym, A. (2020). An adaptive targeted field experiment: Job search assistance for refugees in Jordan. SSRN.

273 Mattos, David Issa, Jan Bosch, and Helena Holmström Olsson. "Multi-armed bandits in the wild: Pitfalls and strategies in online experiments." Information and Software Technology 113 (2019): 68-81.

274 Hopkins, A., Breckon, J., & Lawrence, J. (2020). The experimenter's inventory: a catalogue of experiments for decision-makers and professionals. The Alliance for Useful Evidence.

275 https://www.cecan.ac.uk/wp-content/uploads/2020/08/EPPN-No-03-Agent-Based-Modelling-for-Evaluation.pdf

276 HM Treasury (2020) The Magenta Book: Supplementary Guide: Handling Complexity in Policy Evaluation.

277 Calderoni, F., Campedelli, G. M., Szekely, A., Paolucci, M., & Andrighetto, G. (2022). Recruitment into organized crime: An agent-based approach testing the impact of different policies. Journal of Quantitative Criminology, 38(1), 197-237.

278 Jager, W. (2021). Using agent-based modelling to explore behavioural dynamics affecting our climate. Current opinion in psychology, 42, 133-139.

279 Schlüter, M., et al. (2017). A framework for mapping and comparing behavioural theories in models of social-ecological systems. Ecological Economics, 131, 21-35; Wijermans, N., Boonstra, W. J., Orach, K., Hentati-Sundberg, J., & Schlüter, M. (2020). Behavioural diversity in fishing—Towards a next generation of fishery models. Fish and Fisheries, 21(5), 872-890.

280 Schill, C., et al. (2019). A more dynamic understanding of human behaviour for the Anthropocene. Nature Sustainability, 2(12), 1075-1082.

281 Schlüter, M., et al. (2017).

282 Muelder, H., & Filatova, T. (2018). One theory-many formalizations: Testing different code implementations of the theory of planned behaviour in energy agent-based models. Journal of Artificial Societies and Social Simulation, 21(4).

283 Ewert, B., Loer, K., & Thomann, E. (2020). Beyond nudge: Advancing the state-of-the-art of behavioural public policy and administration. Policy & Politics.

284 Sanbonmatsu, D. M., Cooley, E. H., & Butner, J. E. (2021). The Impact of Complexity on Methods and Findings in Psychological Science. Frontiers in Psychology, 4028.

285 Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. Annual Review of Psychology, 69, 487-510.

286 Sanbonmatsu, D. M., Cooley, E. H., & Butner, J. E. (2021). The Impact of Complexity on Methods and Findings in Psychological Science. Frontiers in Psychology, 4028.

287 Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. Psychological Bulletin, 144(12), 1325.

288 Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716. Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., ... & Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. Nature Human Behaviour, 2(9), 637-644. Sanbonmatsu, D. M., Cooley, E. H., & Butner, J. E. (2021). The Impact of Complexity on Methods and Findings in Psychological Science. Frontiers in Psychology, 4028.

289 Shu, L. L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. Proceedings of the National Academy of Sciences, 109(38), 15197-15200.

290 Behavioural Insights Team, Applying Behavioural Insights to Reduce Fraud, Error and Debt (Cabinet Office, London, 2012) 185, 186. Kettle, S., Hernandez, M., Sanders, M., Hauser, O., & Ruda, S. (2017). Failure to CAPTCHA attention: Null results from an honesty priming experiment in Guatemala. Behavioral Sciences, 7(2), 28.

291 Kristal, A. S., Whillans, A. V., Bazerman, M. H., Gino, F., Shu, L. L., Mazar, N., & Ariely, D. (2020). Signing at the beginning versus at the end does not decrease dishonesty. Proceedings of the National Academy of Sciences, 117(13), 7103-7107. Martuza, J. B., Skard, S. R., Løvlie, L., & Thorbjørnsen, H. (2022). Do honesty-nudges really work? A large-scale field experiment in an insurance context. Journal of Consumer Behaviour.

292 Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. Psychological Science, 22(11), 1359-1366. John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. Psychological science, 23(5), 524-532. Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. Psychological Bulletin, 144(12), 1325. Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. Advances in Methods and Practices in Psychological Science, 3(4), 456-465.

293 Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. Annual Review of Psychology, 69, 511-534.

294 Frias-Navarro, D., Pascual-Llobell, J., Pascual-Soler, M., Perezgonzalez, J., & Berrios-Riquelme, J. (2020). Replication crisis or an opportunity to improve scientific production? European Journal of Education, 55(4), 618-631.

295 McShane, B. B., Böckenholt, U., & Hansen, K. T. (2022). Variation and Covariation in Large-scale Replication Projects: An Evaluation of Replicability. Journal of the American Statistical Association, (Just Accepted), 1-31.

296 Stanley, D. J., & Spence, J. R. (2014). Expectations for replications: Are yours realistic? Perspectives on Psychological Science, 9(3), 305-318.

297 Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. Perspectives on Psychological Science, 9(3), 333-342.

298 Soman, D. & Mazar, N. (2022) The Science of Translation and Scaling. In: Mazar and Soman (eds.) Behavioral Science in the Wild, University of Toronto Press, pp.5-19.

299 McDiarmid, A. D., Tullett, A. M., Whitt, C. M., Vazire, S., Smaldino, P. E., & Stephens, J. E. (2021). Psychologists update their beliefs about effect sizes after replication studies. Nature Human Behaviour, 5(12), 1663-1673.

300 Szaszi, B., Palinkas, A., Palfi, B., Szollosi, A. & Aczel, B. A systematic scoping review of the choice architecture movement: toward understanding when and why nudges work. J. Behav. Decis. Mak. 31, 355–366 (2018). Hummel, D., & Maedche, A. (2019). How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. Journal of Behavioral and Experimental Economics, 80, 47-58. DellaVigna, S., & Linos, E. (2022). RCTs to scale: Comprehensive evidence from two nudge units. Econometrica, 90(1), 81-116.

301 Asendorpf, J. B., et al. (2016). Recommendations for increasing replicability in psychology. Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. Psychological Bulletin, 144(12), 1325.

302 Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. Psychological Bulletin, 144(12), 1325.

303 Allcott, H. (2015). Site selection bias in program evaluation. The Quarterly Journal of Economics, 130(3), 1117-1165. Gerber, A., Huber, G. & Fang, A. Do subtle linguistic interventions priming a social identity as a voter have outsized effects on voter turnout? Evidence from a new replication experiment: outsized turnout effects of subtle linguistic cues. Polit. Psychol. 39, 925–938 (2018). List, J. A. (2022). The voltage effect: How to make good ideas great and great ideas scale. Currency.

304 Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. Nature Human Behaviour, 5(8), 980-989.

305 Klein, R. A., et al. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. Advances in Methods and Practices in Psychological Science, 1(4), 443-490.

306 McShane, B. B., Tackett, J. L., Böckenholt, U., & Gelman, A. (2019). Large-scale replication projects in contemporary psychological research. The American Statistician, 73(sup1), 99-105. Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. Psychological Bulletin, 144(12), 1325. Landy, J. F., et al. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. Psychological Bulletin, 146(5), 451.

307 Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. Psychological bulletin, 144(12), 1325.

308 Sanbonmatsu, D. M., Cooley, E. H., & Butner, J. E. (2021). The Impact of Complexity on Methods and Findings in Psychological Science. Frontiers in Psychology, 4028. McShane, B. B., Tackett, J. L., Böckenholt, U., & Gelman, A. (2019). Large-scale replication projects in contemporary psychological research. The American Statistician, 73(sup1), 99-105.

309 Goodyear, L., Hossain, T., & Soman, D. (2022) Prescriptions for Successfully Scaling Behavioral Interventions. In: Mazar and Soman (eds.) Behavioral Science in the Wild, University of Toronto Press, pp.28-41.

310 For an interesting practice, see: Chorpita, B. F., Becker, K. D., Daleiden, E. L., & Hamilton, J. D. (2007). Understanding the common elements of evidence-based practice: Misconceptions and clinical examples. Journal of the American Academy of Child and Adolescent Psychiatry, 46(5), 647-652.

311 Soman, D. & Mazar, N. (2022) The Science of Translation and Scaling. In: Mazar and Soman (eds.) Behavioral Science in the Wild, University of Toronto Press, pp.5-19.

312 Fiedler, K. (2011). Voodoo correlations are everywhere—not only in neuroscience. Perspectives on psychological science, 6(2), 163-171.

313 Brenninkmeijer, J., Derksen, M., Rietzschel, E., Vazire, S., & Nuijten, M. (2019). Informal laboratory practices in psychology. Collabra: Psychology, 5(1).

314 McShane, B. B., Tackett, J. L., Böckenholt, U., & Gelman, A. (2019). Large-scale replication projects in contemporary psychological research. The American Statistician, 73(sup1), 99-105.

315 Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. Proceedings of the National Academy of Sciences, 113(23), 6454-6459.

316 Landy, J. F. & Crowdsourcing Hypothesis Tests Collaboration. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. Psychological Bulletin, 146(5), 451.

317 Lewin K (1936) Principles of Topological Psychology (McGraw-Hill, New York) trans.

Heider F and Heider G. Camerer CF, Loewenstein G, Rabin M, eds (2011) Advances in Behavioral Economics (Princeton Univ Press, Princeton, NJ).

318 Sanbonmatsu, D. M., Cooley, E. H., & Butner, J. E. (2021). The Impact of Complexity on Methods and Findings in Psychological Science. Frontiers in Psychology, 4028.

319 Gravert, C. (2022) Reminders: Their value and hidden cost. In: Mazar and Soman (eds.) Behavioral Science in the Wild, University of Toronto Press, pp.120-131.

320 Van Ryzin, G. G. (2021). Nudging and Muddling through. Perspectives on Public Management and Governance, 4(4), 339-345.

321 Jang, C., Saldanha, N. A., Singh, A., Adhiambo, J. (2022) Implementing Behavioral Science Insights with Low-income populations in the Global South. In: Mazar and Soman (eds.) Behavioral Science in the Wild, University of Toronto Press, pp.277-283.

322 Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. Psychological bulletin, 144(12), 1325.

323 Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. Nature Human Behaviour, 5(8), 980-989.

324 Choudhary, V., Shunko, M., Netessine, S., & Koo, S. (2021). Nudging drivers to safety: Evidence from a field experiment. Management Science 68 (6), 4196-4214.

325 Bryan, C. J., Tipton, E., & Yeager, D. S. (2021).

326 Bryan, C. J., Tipton, E., & Yeager, D. S. (2021).

327 Allcott, H. (2011). Social norms and energy conservation. Journal of Public Economics, 95(9-10), 1082-1095.

328 Allcott, H. (2015). Site selection bias in program evaluation. The Quarterly Journal of Economics, 130(3), 1117-1165.

329 Sanbonmatsu, D. M., Cooley, E. H., & Butner, J. E. (2021). The Impact of Complexity on Methods and Findings in Psychological Science. Frontiers in Psychology, 4028.

330 Prospect Theory, for example, seems to replicate quite well. Ruggeri, K., et al. (2020). Replicating patterns of prospect theory for decision under risk. Nature Human Behaviour, 4(6), 622-633.

331 Institute for Government and Cabinet Office (2010) MINDSPACE: Influencing behaviour through public policy.

332 Landy, J. F., et al. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. Psychological Bulletin, 146(5), 451. Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. Nature Human Behaviour, 5(8), 980-989.

333 Cartwright, N., & Hardie, J. (2012). Evidence-based policy: A practical guide to doing it better. Oxford University Press. Soman, D. & Mazar, N. (2022) The Science of Translation and Scaling. In: Mazar and Soman (eds.) Behavioral Science in the Wild, University of Toronto Press, pp.5-19.

334 Gelman A (2014) The connection between varying treatment effects and the crisis of unreplicable research: A Bayesian perspective. Journal of Management 41(2):632–643.

335 https://osf.io/zuh93/

336 https://www.nesta.org.uk/blog/mind-gap-between-truth-and-data/

337 https://behavioralscientist.org/breaking-the-silence-can-behavioral-science-confront-structural-racism/

338 Bryan, C. J., Tipton, E., & Yeager, D. S. (2021).

339 McShane, B. B., Tackett, J. L., Böckenholt, U., & Gelman, A. (2019). Large-scale replication projects in contemporary psychological research. The American Statistician, 73(sup1), 99-105.

340 https://youthendowmentfund.org.uk/news/youth-endowment-fund-to-support-grassroots-organisations-to-take-part-in-research-to-find-out-what-works-to-keep-children-safe-from-violence/

341 This point comes from Alex Gyani, BIT's Director of Research & Methodology, APAC.

342 Landy, J. F., et al. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. Psychological Bulletin, 146(5), 451.

343 Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? Behavioral and Brain Sciences, 33(2-3), 61-83.

344 Ruggeri, K., et al. (2020). Replicating patterns of prospect theory for decision under risk. Nature Human Behaviour, 4(6), 622-633.

345 Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. Nature Human Behaviour, 5(8), 980-989.

346 Royer, H. (2021). Benefits of megastudies for testing behavioural interventions. Nature, 8th December 2021.

347 Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. Proceedings of the National Academy of Sciences, 113(23), 6454-6459.

348 Landy, J. F., et al. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. Psychological Bulletin, 146(5), 451.

349 Soman, D. & Mazar, N. (2022) The Science of Translation and Scaling. In: Mazar and Soman (eds.) Behavioral Science in the Wild, University of Toronto Press, pp.5-19 (p.13).

350 List, J. A. (2022). The voltage effect: How to make good ideas great and great ideas scale. Currency.

351 Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implementation science, 4(1), 1-15.

352 Goodyear, L., Hossain, T., & Soman, D. (2022) Prescriptions for Successfully Scaling Behavioral Interventions. In: Mazar and Soman (eds.) Behavioral Science in the Wild, University of Toronto Press, pp.28-41.

353 Landy, J. F., et al. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. Psychological Bulletin, 146(5), 451.

354 Slemrod, J., Blumenthal, M., & Christian, C. (2001). Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota. Journal of Public Economics, 79(3), 455-483.

355 'Betsy Levy Paluck, The Art of Psychology No.1'. Brain Meets World by Behavioral Scientist.

356 De Martino, B., Kumaran, D., Seymour, B. & Dolan, R. J. (2006) Frames, biases, and rational decision-making in the human brain. Science, 313 (5787), 684-687.

357 Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implementation Science, 4(1), 1-15.

358 Mazar and Soman (eds.) Behavioral Science in the Wild. University of Toronto Press

359 Brenninkmeijer, J., Derksen, M., Rietzschel, E., Vazire, S., & Nuijten, M. (2019). Informal laboratory practices in psychology. Collabra: Psychology, 5(1).

360 Kirsh, S. R., Lawrence, R. H., & Aron, D. C. (2008). Tailoring an intervention to the context and system redesign related to the intervention: A case study of implementing shared medical appointments for diabetes. Implementation Science, 3(1), 1-15. Goswami, I. & Urminsky, O. (2022) Why Many Behavioral Interventions Have Unpredictable Effects in the Wild: The Conflicting Consequences Problem. In: Mazar and Soman (eds.) Behavioral Science in the Wild, University of Toronto Press, pp.65-81.

361 Hallsworth, M., Chadborn, T., Sallis, A., Sanders, M., Berry, D., Greaves, F., & Davies, S. C. (2016). Provision of social norm feedback to high prescribers of antibiotics in general practice: a pragmatic national randomised controlled trial. The Lancet, 387(10029), 1743-1752.

362 https://www1.racgp.org.au/newsgp/professional/nudge-letters-prompt-sustained-drop-in-antibiotic; Schwartz, K. L., et al. (2021). Effect of antibiotic-prescribing feedback to high-volume primary care physicians on number of antibiotic prescriptions: a randomized clinical trial. JAMA Internal Medicine, 181(9), 1165-1173. https://www.ars.toscana.it/2-articoli/4134-nudge-per-uso-prudente-antibiotici-in-toscana-intervento-su-medici-di-famiglia.html#; Bradley, D. T., Allen, S. E., Quinn, H., Bradley, B., & Dolan, M. (2019). Social norm feedback reduces primary care antibiotic prescribing in a regression discontinuity study. Journal of Antimicrobial Chemotherapy, 74(9), 2797-2802.

363 https://www.bi.team/blogs/amr-blog/

364 McShane, B. B., Tackett, J. L., Böckenholt, U., & Gelman, A. (2019). Large-scale replication projects in contemporary psychological research. The American Statistician, 73(sup1), 99-105.

365 Fried, E. I. (2020). Lack of theory building and testing impedes progress in the factor and network literature. Psychological Inquiry, 31(4), 271-288.

366 Bryan, C.J., Tipton, E. & Yeager, D.S. Behavioural science is unlikely to change the world without a heterogeneity revolution. Nat Hum Behav 5, 980–989 (2021).

367 McShane, B. B., Böckenholt, U., & Hansen, K. T. (2022). Variation and Covariation in Large-scale Replication Projects: An Evaluation of Replicability. Journal of the American Statistical Association, (just-accepted), 1-31.

368 Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. Psychonomic bulletin & review, 26(5), 1596-1618. Borsboom, D., van der Maas, H. L., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. Perspectives on Psychological Science, 16(4), 756-766.

369 Sanbonmatsu, D. M., & Johnston, W. A. (2019). Redefining science: The impact of complexity on theory development in social and behavioral research. Perspectives on Psychological Science, 14(4), 672-690.

370 Sanbonmatsu, D. M., Cooley, E. H., & Butner, J. E. (2021). The Impact of Complexity on Methods and Findings in Psychological Science. Frontiers in Psychology, 4028.

371 Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. Psychonomic Bulletin & Review, 26(5), 1596-1618.

372 Fried, E. I. (2020). Theories and models: What they are, what they are for, and what they are about. Psychological Inquiry, 31(4), 336-344.

373 Maatman, F. O. (2021). Psychology's Theory Crisis, and Why Formal Modelling Cannot Solve It.

374 Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. Psychonomic Bulletin & Review, 26(5), 1596-1618.

375 Hallsworth, M. & Kirkman, E. (2020) Behavioral Insights. MIT Press.

376 Pennycook, G. (2017). A perspective on the theoretical foundation of dual process models. In De Neys, W. (Ed.), Dual Process Theory 2.0 (pp. 13–35). De Neys, W. (2022) Advancing theorizing about fast-and-slow thinking. In press at Behavioral and Brain Sciences. Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. Perspectives on psychological science, 4(6), 533-550.

377 Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. American Economic Review, 105(5), 1–33. Kosters, M., & Van der Heijden, J. (2015). From mechanism to virtue: Evaluating Nudge theory. Evaluation, 21(3), 276-291. Callaway, F., Hardy, M., & Griffiths, T. (2022). Optimal nudging for cognitively bounded agents: A framework for modeling, predicting, and controlling the effects of choice architectures.

378 Muthukrishna, M., & Henrich, J. (2019). A problem in theory. Nature Human Behaviour, 3(3), 221-229.

379 https://www.psychologicalscience.org/observer/grand-challenges; Schlüter, M., et al. (2017). A framework for mapping and comparing behavioural theories in models of social-ecological systems. Ecological Economics, 131, 21-35.

380 Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. Perspectives on Psychological Science, 12(6), 1100-1122.

381 Muthukrishna, M., & Henrich, J. (2019). A problem in theory. Nature Human Behaviour, 3(3), 221-229.

382 Muthukrishna, M., & Henrich, J. (2019).

383 https://www.worksinprogress.co/issue/biases-the-wrong-model/

384 It may be possible to reconcile the two - i.e., that negativity bias is processed more as negative outcomes relating to others, rather than oneself, but this requires a deeper level of analysis than is usually present. Sharot, T. (2011). The optimism bias. Current biology, 21(23), R941-R945. Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. Personality and Social Psychology Review, 5(4), 296-320.

385 Rizzo, M. J., & Whitman, G. (2023). The unsolved Hayekian knowledge problem in behavioral economics. Behavioural Public Policy, 7(1), 199-211.

386 Schimmelpfennig, R., & Muthukrishna, M. (2023). Cultural evolutionary behavioural science in public policy. Behavioural Public Policy, 1-31. doi:10.1017/bpp.2022.40. Kwan, V. S., John, O. P., Kenny, D. A., Bond, M. H., & Robins, R. W. (2004). Reconceptualizing individual differences in self-enhancement bias: an interpersonal approach. Psychological Review, 111(1), 94. Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. Psychological Bulletin, 130(5), 711.

387 http://behavioralscientist.org/there-is-more-to-behavioral-science-than-biases-and-fallacies/

388 Muthukrishna, M., & Henrich, J. (2019). A problem in theory. Nature Human Behaviour, 3(3), 221-229.

389 Rand, D. G. (2016). Cooperation, fast and slow: Meta-analytic evidence for a theory of social heuristics and self-interested deliberation. Psychological science, 27(9), 1192-1206. Gelfand, M. J. (2018). Rule makers, rule breakers: how culture wires our minds, shapes our nations and drive our differences. Robinson.

390 Manzi, J. (2012). Uncontrolled: The surprising payoff of trial-and-error for business, politics, and society. Basic Books (AZ).

391 https://www.theguardian.com/technology/2022/jan/09/are-we-witnessing-the-dawn-of-post-theory-science

392 Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. Perspectives on Psychological Science, 12(6), 1100-1122.

393 Van Ryzin, G. G. (2021). Nudging and Muddling through. Perspectives on Public Management and Governance, 4(4), 339-345.

394 Van Ryzin, G. G. (2021).

395 Schmidt, R., & Stenger, K. (2021). Behavioral brittleness: The case for strategic behavioral public policy. Behavioural Public Policy, 1-26.

396 Yeung, K. (2012). Nudge as fudge. Mod. L. Rev., 75, 122.

397 van Rooij, I., & Baggio, G. (2021). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. Perspectives on Psychological Science, 16(4), 682-697.

398 Muthukrishna, M., & Henrich, J. (2019). A problem in theory. Nature Human Behaviour, 3(3), 221-229.

399 Schimmelpfennig, R., & Muthukrishna, M. (2023). Cultural evolutionary behavioural science in public policy. Behavioural Public Policy, 1-31. doi:10.1017/bpp.2022.40. Smaldino, P. E. (2020). How to build a strong theoretical foundation. Psychological Inquiry, 31(4), 297-301.

400 Fried, E. I. (2020). Theories and models: What they are, what they are for, and what they are about. Psychological Inquiry, 31(4), 336-344.

401 Sanbonmatsu, D. M., & Johnston, W. A. (2019). Redefining science: The impact of complexity on theory development in social and behavioral research. Perspectives on Psychological Science, 14(4), 672-690. Smaldino, P. (2019). Better methods can't make up for mediocre theory. Nature, 575(7783), 9-10.

402 van Rooij, I., & Baggio, G. (2021). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. Perspectives on Psychological Science, 16(4), 682-697. Borsboom, D., van der Maas, H. L., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. Perspectives on Psychological Science, 16(4), 756-766. Smaldino, P. E. (2020). How to build a strong theoretical foundation. Psychological Inquiry, 31(4), 297-301.

403 Schimmelpfennig, R., & Muthukrishna, M. (2023).

404 This list brings together aspects of two separate but related ideas: "middle-range theories" and "practical theories" in a composite that we think is helpful. erton, R. K., & Merton, R. C. (1968). Social theory and social structure. Simon and Schuster. Berkman, E. T., & Wilson, S. M. (2021). So useful as a good theory? The practicality crisis in (social) psychological theory. Perspectives on psychological science, 16(4), 864-874.

405 Abner, G. B., Kim, S. Y., & Perry, J. L. (2017). Building evidence for public human resource management: Using middle range theory to link theory and data. Review of Public Personnel Administration, 37(2), 139-159.

406 Moore, L. F., Johns, G., & Pinder, C. C. (1980). Toward middle range theory. Middle range theory and the study of organizations, 1-16.

407 Sanbonmatsu, D. M., Cooley, E. H., & Butner, J. E. (2021). The Impact of Complexity on Methods and Findings in Psychological Science. Frontiers in Psychology, 4028.

408 Berkman, E. T., & Wilson, S. M. (2021). So useful as a good theory? The practicality crisis in (social) psychological theory. Perspectives on Psychological Science, 16(4), 864-874.

409 Lieder, F., & Griffiths, T. L. (2020a). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. Behavioral and Brain Sciences, 43. doi:10.1017/S0140525X1900061X

410 Callaway, F., Hardy, M., & Griffiths, T. (2022). Optimal nudging for cognitively bounded agents: A framework for modeling, predicting, and controlling the effects of choice architectures. Working Paper.

411 Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. Science, 185 (4157), 1124-1131.

412 Lieder, F., Griffiths, T. L., M Huys, Q. J., & Goodman, N. D. (2018). The anchoring bias reflects rational use of cognitive resources. Psychonomic Bulletin & Review, 25 (1), 322-349.

413 Lieder, F., Griffiths, T. L., M Huys, Q. J., & Goodman, N. D. (2018).

414 Lieder, F., & Griffiths, T. L. (2020a).

415 Lieder, F., & Griffiths, T. L. (2020a).

416 Lieder, F., & Griffiths, T. L. (2020a).

417 The following draws on Callaway, F., Hardy, M., & Griffiths, T. (2022).

418 Lieder, F., & Griffiths, T. L. (2020b). Advancing rational analysis to the algorithmic level. Behavioral and Brain Sciences, 43.

419 Atmanspacher, H., Basieva, I., Busemeyer, J. R., Khrennikov, A. Y., Pothos, E. M., Shiffrin, R. M., & Wang, Z. (2020). What are the appropriate axioms of rationality for reasoning under uncertainty with resource-constrained systems? Behavioral and Brain Sciences, 43.

420 Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. Scientometrics 90, 891–904. Uchino, B. N., Thoman, D., & Byerly, S. (2010). Inference patterns in theoretical social psychology: Looking back as we move forward. Social and Personality Psychology Compass, 4(6), 417-427. Sanbonmatsu, D. M., Posavac, S. S., Behrends, A. A., Moore, S. M., & Uchino, B. N. (2015). Why a confirmation strategy dominates psychological science. PloS One, 10(9), e0138197.

421 Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. Perspectives on Psychological Science, 7(5), 411-426.

422 Roese, N. J., & Vohs, K. D. (2012).

423 Roese, N. J., & Vohs, K. D. (2012).

424 Munnich, E., & Ranney, M. A. (2019). Learning from surprise: Harnessing a metacognitive surprise signal to build and adapt belief networks. Topics in Cognitive Science, 11(1), 164-177. Pezzo, M. (2003). Surprise, defence, or making sense: What removes hindsight bias? Memory, 11(4-5), 421-441. Calvillo, D. P., & Gomes, D. M. (2011). Surprise influences hindsight–foresight differences in temporal judgments of animated automobile accidents. Psychonomic Bulletin & Review, 18(2), 385-391.

425 Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: a failure to disagree. American Psychologist, 64(6), 515. Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self-evaluations undermine students' learning and retention. Learning and Instruction, 22(4), 271-280.

426 Henriksen, K., & Kaplan, H. (2003). Hindsight bias, outcome knowledge and adaptive learning. BMJ Quality & Safety, 12(suppl 2), ii46-ii50. Bernstein, D. M., Aßfalg, A., Kumar, R., & Ackerman, R. (2016). Looking backward and forward on hindsight bias. In: J. Dunlosky & S. K. Tauber (eds.), The Oxford Handbook of Metamemory (pp. 289–304). Oxford University Press.

427 Bukszar, E., & Connolly, T. (1988). Hindsight bias and strategic choice: Some problems in learning from experience. Academy of Management Journal, 31(3), 628-641.

428 DellaVigna, S., Pope, D., & Vivalt, E. (2019). Predict science to improve science. Science, 366(6464), 428-429.

429 DellaVigna, S., Pope, D., & Vivalt, E. (2019). Predict science to improve science. Science, 366(6464), 428-429.

430 Munnich, E., & Ranney, M. A. (2019). Learning from surprise: Harnessing a metacognitive surprise signal to build and adapt belief networks. Topics in Cognitive Science, 11(1), 164-177.

431 Munnich, E., & Ranney, M. A. (2019). Learning from surprise: Harnessing a metacognitive surprise signal to build and adapt belief networks. Topics in Cognitive Science, 11(1), 164-177.

432 https://www.bi.team/blogs/how-government-can-predict-the-future/ ; https://www.bi.team/blogs/are-you-well-calibrated-results-from-a-survey-of-1154-bit-readers/

433 Deshpande, M., & Dizon-Ross, R. (2022). The (lack of) anticipatory effects of the social safety net on human capital investment. Working Paper.

434 DellaVigna, S., Pope, D., & Vivalt, E. (2019). Predict science to improve science. Science, 366(6464), 428-429.

435 DellaVigna, S., & Pope, D. (2018). Predicting experimental results: who knows what? Journal of Political Economy, 126(6), 2410-2456. DellaVigna, S., & Pope, D. (2018). What motivates effort? Evidence and expert forecasts. The Review of Economic Studies, 85(2), 1029-1069. Otis, N. G. (2021). Forecasting in the Field. Working Paper.

436 Abaluck, J., et al. (2021). Impact of community masking on COVID-19: A cluster-randomized trial in Bangladesh. Science, eabi9069. Dimant, E., Clemente, E. G., Pieper, D., Dreber, A., & Gelfand, M. (2022). Politicizing mask-wearing: predicting the success of behavioral interventions among republicans and democrats in the US. Scientific Reports, 12(1), 1-12. Bowen, D. (2022) Simple models predict behavior at least as well as behavioral scientists. Working Paper.

437 DellaVigna, S., & Pope, D. (2018). Predicting experimental results: who knows what? Journal of Political Economy, 126(6), 2410-2456. DellaVigna, S., & Pope, D. (2018). What motivates effort? Evidence and expert forecasts. Dimant, E., Clemente, E. G., Pieper, D., Dreber, A., & Gelfand, M. (2022). Politicizing mask-wearing: predicting the success of behavioral interventions among republicans and democrats in the US. Scientific Reports, 12(1), 1-12. Otis, N. G. (2021). Forecasting in the Field. Working Paper. Counterexample: Milkman, K., et al. (2022). A 680,000-person megastudy of nudges to encourage vaccination in pharmacies. Proceedings of the National Academy of Sciences, 119(6), e2115126119.

438 DellaVigna, S., & Linos, E. (2022). RCTs to scale: Comprehensive evidence from two nudge units. Econometrica, 90(1), 81-116.

439 Otis, N. G. (2021). Forecasting in the Field. Working Paper.

440 Pekrun, R., & Stephens, E. J. (2012). Academic emotions. In: APA Educational Psychology Handbook, Vol 2: Individual differences and cultural and contextual factors. (pp. 3-31). American Psychological Association.

441 Ackerman, R., Bernstein, D. M., & Kumar, R. (2020). Metacognitive hindsight bias. Memory & Cognition, 48(5), 731-744.

442 DellaVigna, S., Pope, D., & Vivalt, E. (2019). Predict science to improve science. Science, 366(6464), 428-429.

443 Nerantzaki, K., Efklides, A., & Metallidou, P. (2021). Epistemic emotions: Cognitive underpinnings and relations with metacognitive feelings. New Ideas in Psychology, 63, 100904.

444 Munnich, E., & Ranney, M. A. (2019). Learning from surprise: Harnessing a metacognitive surprise signal to build and adapt belief networks. Topics in Cognitive Science, 11(1), 164-177.

445 Speirs-Bridge, A., Fidler, F., McBride, M., Flander, L., Cumming, G., & Burgman, M. (2010). Reducing overconfidence in the interval judgments of experts. Risk Analysis, 30: 512-523.; Angner, E. (2006). Economists as experts: Overconfidence in theory and practice. Journal of Economic Methodology, 13(1), 1-24.

446 Byrne, D., & Callaghan, G. (2013). Complexity theory and the social sciences: The state of the art. Routledge.

447 Liscow, Z. & Markovits (2022) Democratizing Behavioral Economics. Working Paper.

448 Rizzo, M. J., & Whitman, G. (2020). Escaping paternalism: Rationality, behavioral economics, and public policy. Cambridge University Press.

449 Ruggeri, K., Panin, A., Vdovic, M. et al. (2022). The globalizability of temporal discounting. Nature Human Behaviour. https://doi.org/10.1038/s41562-022-01392-w

450 Dorison, C. A., & Heller, B. H. Observers penalize decision makers whose risk preferences are unaffected by loss-gain framing. Journal of Experimental Psychology: General, https://doi.org/10.1037/xge0001187

451 Hallsworth, M. & Kirkman, E. (2020) Behavioral Insights. MIT Press.

452 Smaldino, P. E. (2020). How to build a strong theoretical foundation. Psychological Inquiry, 31(4), 297-301. Lamont, M., Adler, L., Park, B. Y., & Xiang, X. (2017). Bridging cultural sociology and cognitive psychology in three contemporary research programmes. Nature Human Behaviour, 1(12), 866-872.

453 Schmidt, R., & Stenger, K. (2021). Behavioral brittleness: the case for strategic behavioral public policy. Behavioural Public Policy, 1-26.

454 Resource-rational analysis, which I discuss later, provides new evidence that 'When people seem to act against their own interests, they may be doing so in light of priorities and constraints that aren't obvious to outsiders – or even to themselves.' Cowles, H. M., & Kreiner, J. (2020). Another claim for cognitive history. Behavioral and Brain Sciences, 43.

455 https://www.bi.team/blogs/green-means-go-how-to-help-patients-make-informed-choices-about-their-healthcare/

456 Elster, J. (2008). Reason and rationality. Princeton University Press. Pinker, S. (2021). Rationality.

457 Commentators like Steven Pinker argue that rationality is based on 'epistemic humility' because a commitment to rationality does not imply that you have access to objective truth, but rather that you support effective tools for getting closer to the truth. But dealing with disagreement is a crucial way that these tools work and delegitimizing opposing stances as 'irrational' prevents that from happening.

458 Daniel Kahneman states that 'I often cringe when my work with Amos [Tversky] is credited with demonstrating that human choices are irrational, when in fact our research only showed that Humans are not well described by the rational-agent model.' Thinking, Fast and Slow, p.411.

459 https://behavioralscientist.org/epistemic-humility-coronavirus-knowing-your-limits-in-a-pandemic/ - see also Porter, T., Elnakouri, A., Meyers, E. A., Shibayama, T., Jayawickreme, E., & Grossmann, I. (2022). Predictors and consequences of intellectual humility. Nature Reviews Psychology, 1-13.

460 Mazzocchi, F. (2021). Drawing lessons from the COVID-19 pandemic: science and epistemic humility should go together. History and Philosophy of the Life Sciences, 43(3), 1-5.

461 Behavioural Insights Team (2018) Behavioural Government: Using behavioural science to improve how governments make decisions.

462 Van Bavel, R., & Dessart, F. J. (2018). The case for qualitative methods in behavioural studies for EU policy-making. Publications Office of the European Union: Luxembourg. https://www.povertyactionlab.org/blog/8-11-21/strengthening-randomized-evaluations-through-incorporating-qualitative-research-part-1

463 https://www.bi.team/blogs/citizens-jury/

464 Schmidt, R., & Stenger, K. (2021). Behavioral brittleness: the case for strategic behavioral public policy. Behavioural Public Policy, 1-26.

465 Walton, G. M., & Wilson, T. D. (2018). Wise interventions: Psychological remedies for social and personal problems. Psychological review, 125(5), 617.

466 See the emphasis in Institute for Government and Cabinet Office (2010) MINDSPACE: Influencing behaviour through public policy.

467 Lewis Jr, N. A. (2021). What counts as good science? How the battle for methodological legitimacy affects public psychology. American Psychologist, 76.

468 https://behavioralscientist.org/breaking-the-silence-can-behavioral-science-confront-structural-racism/

469 https://behavioralscientist.org/breaking-the-silence-can-behavioral-science-confront-structural-racism/

470 Schill, C., et al. (2019). A more dynamic understanding of human behaviour for the Anthropocene. Nature Sustainability, 2(12), 1075-1082. Lewis Jr, N. A. (2021).

471 Hallsworth, M. & Kirkman, E. (2020), pp.176-179.

472 DiMaggio, P. Culture and cognition. Annu. Rev. Sociol. 23, 263–287 (1997). Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: holistic versus analytic cognition. Psychological Review, 108(2), 291. Vaisey, S. (2009). Motivation and justification: A dual-process model of culture in action. American Journal of Sociology, 114(6), 1675-1715.

473 Lamont, M., Adler, L., Park, B. Y., & Xiang, X. (2017). Bridging cultural sociology and cognitive psychology in three contemporary research programmes. Nature Human Behaviour, 1(12), 866-872.

474 Wang, Q. (2021). The cultural foundation of human memory. Annual review of Psychology, 72, 151-179.

475 Hogg, M. A., & Reid, S. A. (2006). Social identity, self-categorization, and the communication of group norms. Communication Theory, 16(1), 7-30.

476 Leggett, W. (2014). The politics of behaviour change: Nudge, neoliberalism and the state. Policy & Politics, 42(1), 3-19. DiMaggio, P., & Markus, H. R. (2010). Culture and social psychology: Converging perspectives. Social Psychology Quarterly, 73(4), 347-352.

477 Swidler, A. (1986). Culture in action: Symbols and strategies. American Sociological Review, 273-286.

478 Mullainathan, S., & Shafir, E. (2013). Scarcity: Why having too little means so much. Macmillan.

479 World Bank (2014). World development report 2015: Mind, society, and behavior. The World Bank.

480 We take this example and the argument below from Lamont, M., Adler, L., Park, B. Y., & Xiang, X. (2017). Bridging cultural sociology and cognitive psychology in three contemporary research programmes. Nature Human Behaviour, 1(12), 866-872.

481 Lamont, M., Adler, L., Park, B. Y., & Xiang, X. (2017).

482 Lamont, M., Adler, L., Park, B. Y., & Xiang, X. (2017).

483 Richardson, L., & John, P. (2021). Co-designing behavioural public policy: lessons from the field about how to 'nudge plus'. Evidence & Policy, 17(3), 405-422.

484 Grüne-Yanoff, T. (2012). Old wine in new casks: libertarian paternalism still violates liberal principles. Social Choice and Welfare, 38(4), 635-645. Rizzo, M. J., & Whitman, G. (2020). Escaping paternalism: Rationality, behavioral economics, and public policy. Cambridge University Press.

485 See Chapter 5 of Hallsworth, M. & Kirkman, E. (2020).

486 John, P., et al. (2013). Nudge, nudge, think, think: Experimenting with ways to change civic behaviour. A&C Black.

487 Banerjee, S., & John, P. (2020). Nudge plus: incorporating reflection into behavioral public policy. Behavioural Public Policy, 1-16.

488 Reijula, S., & Hertwig, R. (2022). Self-nudging and the citizen choice architect. Behavioural Public Policy, 6(1), 119-149.

489 Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. Perspectives on Psychological Science, 12(6), 973-986.

490 https://www.thersa.org/reports/steer-the-report

491 In the figure, nudges are partly present in the 'reflection' row. This is because 'many nudges do require citizens to think and indeed reflect', and therefore overlap with nudge pluses. Banerjee, S., & John, P. (2020). Nudge plus: incorporating reflection into behavioral public policy. Behavioural Public Policy, 1-16.

492 Einfeld, C., & Blomkamp, E. (2021). Nudge and co-design: complementary or contradictory approaches to policy innovation? Policy Studies, 1-19.

493 Richardson, L., & John, P. (2021). Co-designing behavioural public policy: lessons from the field about how to 'nudge plus'. Evidence & Policy, 17(3), 405-422.

494 Einfeld, C., & Blomkamp, E. (2021).

495 Hills, John (2007) Pensions, public opinion and policy. In: Hills, John, Le Grand, Julian and Piachaud, David, (eds.) Making Social Policy Work. CASE studies on poverty, place and policy. The Policy Press, Bristol, UK, pp. 221-243. ISBN 9781861349576

496 This is similar to Cass Sunstein's arguments about 'choosing not to choose'. Sunstein, C. R. (2015). Choosing not to choose: Understanding the value of choice. Oxford University Press, USA.

497 Reijula, S., & Hertwig, R. (2022). Self-nudging and the citizen choice architect. Behavioural Public Policy, 6(1), 119-149.

498 Moreover, policy makers will still have decided to invest scarce resources in a boost intervention, which may mean other options are not available. Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. Perspectives on Psychological Science, 12(6), 973-986.Grüne-Yanoff, T., Marchionni, C., & Feufel, M. A. (2018). Toward a framework for selecting behavioural policies: How to choose between boosts and nudges. Economics & Philosophy, 34(2), 243-266.

499 Part of the vision of boosting is that people may develop the ability to transfer heuristics from the original site of learning to new areas.

500 Sims and Muller (2018) argue that claims such as 'nudges are illiberal, but boosts are not' are not valid. Sims, A., & Müller, T. M. (2019). Nudge versus boost: A distinction without a normative difference. Economics & Philosophy, 35(2), 195-222.

501 Miller, G. A. (1969). Psychology as a means of promoting human welfare. American Psychologist, 24(12), 1063.

502 Thaler, R. & Sunstein, C. (2008), p.252.

503 Banerjee, S., & John, P. (2020). Nudge plus: incorporating reflection into behavioral public policy. Behavioural Public Policy, 1-16.

504 Hallsworth, M. & Kirkman, E. (2020).

505 Einfeld, C., & Blomkamp, E. (2021). Nudge and co-design: complementary or contradictory approaches to policy innovation? Policy Studies, 1-19.

506 Bason, C. (2018). Leading public sector innovation: Co-creating for a better society. Policy Press, pp.64–65.

507 Strassheim, H. (2020). De-biasing democracy. Behavioural public policy and the post-democratic turn. Democratization, 27(3), 461-476.

508 Rouyard, T., Engelen, B., Papanikitas, A., & Nakamura, R. (2022). Boosting healthier choices. BMJ, 376.

509 Ralph Hertwig offers a different set of criteria in Hertwig, R. (2017). When to consider boosting: some rules for policy-makers. Behavioural Public Policy, 1(2), 143-161. See also: Grüne-Yanoff, T., Marchionni, C., & Feufel, M. A. (2018). Toward a framework for selecting behavioural policies: How to choose between boosts and nudges. Economics & Philosophy, 34(2), 243-266.

510 Hertwig, R. (2017). When to consider boosting: some rules for policy-makers. Behavioural Public Policy, 1(2), 143-161.

511 Kristal and Santos usefully distinguish between 'encapsulated' and 'attentional' biases; the former are much more difficult to de-bias. Kristal, A. S., & Santos, L. R. (2021). GI Joe Phenomena: Understanding the Limits of Metacognitive Awareness on Debiasing.

512 Hertwig, R. (2017).

513 Weimer, D. L. (2020). When are nudges desirable? Benefit validity when preferences are not consistently revealed. Public Administration Review, 80(1), 118-126.; Hertwig, R. (2017).

514 Note that this criterion deals with someone's view on the choices available; the preceding one deals with whether they want to engage with the choice at all.

515 Hertwig, R. (2017).

516 Hertwig, R. (2017).

517 Hertwig, R. (2017).

518 Mrkva, K., Posner, N. A., Reeck, C., & Johnson, E. J. (2021). Do nudges reduce disparities? Choice architecture compensates for low consumer knowledge. Journal of Marketing, 0022242921993186.

519 Banerjee, S., & John, P. (2020). Nudge plus: incorporating reflection into behavioral public policy. Behavioural Public Policy, 1-16. Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. Perspectives on Psychological Science, 12(6), 973-986.

520 Bradt, J. (2019). Comparing the effects of behaviorally informed interventions on flood insurance demand: an experimental analysis of 'boosts' and 'nudges'. Behavioural Public Policy, 1-31.

521 Folke, T., Bertoldo, G., D'Souza, D., Alì, S., Stablum, F., & Ruggeri, K. (2021). Boosting promotes advantageous risk-taking. Humanities and Social Sciences Communications, 8(1), 1-10.

522 Mühlböck, M., Kalleitner, F., Steiber, N., & Kittel, B. (2022). Information, reflection, and successful job search: A labor market policy experiment. Social Policy & Administration, 56(1), 48-72.

523 Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. Proceedings of the National Academy of Sciences, 114(48), 12714-12719.

524 'Big-data studies of human behaviour need a common language.' Nature July 8, 2021. Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. Perspectives on Psychological Science, 12(6), 1100-1122.

525 Athey, S. (2018). The impact of machine learning on economics. In The economics of artificial intelligence: An agenda (pp. 507-547). University of Chicago Press.

526 Smith, W. R. (1956). Product differentiation and market segmentation as alternative marketing strategies. Journal of Marketing, 21(1), 3-8.

527 Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. Nature Human Behaviour, 5(8), 980-989.

528 Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 113(523), 1228-1242. Künzel, S. R., Sekhon, J. S., Bickel, P. J., & Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. Proceedings of the National Academy of Sciences, 116(10), 4156-4165.

529 Todd-Blick, A., Spurlock, C. A., Jin, L., Cappers, P., Borgeson, S., Fredman, D., & Zuboy, J. (2020). Winners are not keepers: Characterizing household engagement, gains, and energy patterns in demand response using machine learning in the United States. Energy Research & Social Science, 70, 101595.

530 Ribers, M. A., & Ullrich, H. (2022). Machine learning and physician prescribing: a path to reduced antibiotic use. Working paper.

531 Mills, S. (2022). Personalized nudging. Behavioural Public Policy, 6(1), 150-159.

532 Mohlmann, M. (2021) Algorithmic Nudges Don't Have to Be Unethical. Harvard Business Review. Mills, S. (2022). Personalized nudging. Behavioural Public Policy, 6(1), 150-159.

533 Morozovaite, V. (2021). Two sides of the digital advertising coin: putting hypernudging into perspective. Mkt. & Competition L. Rev., 5, 105. Teeny, J. D., Siev, J. J., Briñol, P., & Petty, R. E. (2021). A review and conceptual framework for understanding personalized matching effects in persuasion. Journal of Consumer Psychology, 31(2), 382-414.

534 https://www.research-live.com/article/opinion/new-frontiers-the-growth-of-bespoke-nudging/id/5066852

535 Soman, D. & Mazar, N. (eds) Behavioral Science in the Wild, University of Toronto Press, pp.284-291; https://www.aitimejournal.com/interview-with-ganna-pogrebna-lead-for-behavioral-data-science-alan-turing-institute

536 Teeny, J. D., Siev, J. J., Briñol, P., & Petty, R. E. (2021). A review and conceptual framework for understanding personalized matching effects in persuasion. Journal of Consumer Psychology, 31(2), 382-414.

537 Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. Proceedings of the national academy of sciences, 114(48), 12714-12719.

538 Hullman, J., Kapoor, S., Nanayakkara, P., Gelman, A., & Narayanan, A. (2022). The worst of both worlds: A comparative analysis of errors in learning from data in psychology and machine learning. arXiv preprint arXiv:2203.06498. McDermott, M. B., Wang, S., Marinsek, N., Ranganath, R., Foschini, L., & Ghassemi, M. (2021). Reproducibility in machine learning for health research: Still a ways to go. Science Translational Medicine, 13(586), eabb1655.

539 https://www.nature.com/articles/d41586-018-05707-8

540 Salganik, M. J., et al. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. Proceedings of the National Academy of Sciences, 117(15), 8398-8403.

541 Peer, E., Egelman, S., Harbach, M., Malkin, N., Mathur, A., & Frik, A. (2020). Nudge me right: Personalizing online security nudges to people's decision-making styles. Computers in Human Behavior, 109, 106347.

542 https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/; https://thenewstack.io/clean-data-is-the-foundation-of-effective-machine-learning/

543 Mills, S. (2022). Personalized nudging. Behavioural Public Policy, 6(1), 150-159.

544 https://hdsr.mitpress.mit.edu/pub/4cg8dhgr/release/3

545 Ruggeri, K., Benzerga, A., Verra, S., & Folke, T. (2020). A behavioral approach to personalizing public health. Behavioural Public Policy, 1-13.

546 Horkheimer, M. & Adorno, T. (1972). Dialectic of Enlightenment (org. pub. 1947).

547 Mills, S. (2022). Personalized nudging. Behavioural Public Policy, 6(1), 150-159.

548 Mohlmann, M. (2021) Algorithmic Nudges Don't Have to Be Unethical. Harvard Business Review.

549 Summers, C. A., Smith, R. W., & Reczek, R. W. (2016). An audience of one: Behaviorally targeted ads as implied social labels. Journal of Consumer Research, 43(1), 156-178.

550 Spencer, S. B. (2020). The problem of online manipulation. U. Ill. L. Rev., 959.

551 Susser, D., Roessler, B., & Nissenbaum, H. (2019). Online manipulation: Hidden influences in a digital world. Geo. L. Tech. Rev., 4, 1.

552 Morozovaite, V. (2021). Two sides of the digital advertising coin: putting hypernudging into perspective. Mkt. & Competition L. Rev., 5, 105.

553 Mohsen Abbasi, Sorelle A Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian. 2019. Fairness in representation: quantifying stereotyping as a representational harm. In: Proc. of the 2019 SIAM International Conference on Data Mining. SIAM, 801–809

554 Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453. Bigman, Y., Gray, K., Waytz, A., Arnestad, M., & Wilson, D. (2020). Algorithmic discrimination causes less moral outrage than human discrimination. In Press at Journal of Experimental Psychology: General. Athey, S. (2018). The impact of machine learning on economics. In The economics of artificial intelligence: An agenda (pp. 507-547). University of Chicago Press.

555 National Audit Office (2022) Report on Accounts: Department for Work and Pensions.

556 Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

557 Kozyreva, A., Lorenz-Spreen, P., Hertwig, R., Lewandowsky, S., & Herzog, S. M. (2021). Public attitudes towards algorithmic personalization and use of personal data online: Evidence from Germany, Great Britain, and the United States. Humanities and Social Sciences Communications, 8(1), 1-11.

558 Teeny, J. D., Siev, J. J., Briñol, P., & Petty, R. E. (2021). A review and conceptual framework for understanding personalized matching effects in persuasion. Journal of Consumer Psychology, 31(2), 382-414.

559 De Jonge, P., Verlegh, P. & Zeelenberg, M. (2022) If You Want People to Accept Your Intervention, Don't Be Creepy. In: Soman, D. & Mazar, N. (eds) Behavioral Science in the Wild, University of Toronto Press, pp.284-291.

560 van Doorn & Hoekstra, 2013; Kim net al., 2019a; White, T. B., Zahay, D. L., Thorbjørnsen, H., & Shavitt, S. (2008). Getting too personal: Reactance to highly personalized email solicitations. Marketing Letters, 19(1), 39-50.

561 Briñol, P., Rucker, D. D., & Petty, R. E. (2015). Naïve theories about persuasion: Implications for information processing and consumer attitude change. International Journal of Advertising, 34(1), 85-106. Reinhart, A. M., Marshall, H. M., Feeley, T. H., & Tutzauer, F. (2007). The persuasive effects of message framing in organ donation: The mediating role of psychological reactance. Communication Monographs, 74(2), 229-255.

562 Clark, J. K., Wegener, D. T., & Fabrigar, L. R. (2008). Attitude accessibility and message processing: The moderating role of message position. Journal of Experimental Social Psychology, 44(2), 354-361.

563 Derricks, V., & Earl, A. (2019). Information targeting increases the weight of stigma: Leveraging relevance backfires when people feel judged. Journal of Experimental Social Psychology, 82, 277-293.; White, K., & Argo, J. J. (2009). Social identity threat and consumer preferences. Journal of Consumer Psychology, 19(3), 313-325.

564 Derricks, V., & Earl, A. (2019). Information targeting increases the weight of stigma: Leveraging relevance backfires when people feel judged. Journal of Experimental Social Psychology, 82, 277-293.

565 Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. Proceedings of the National Academy of Sciences, 114(48), 12714-12719.

566 Pierson, E., Cutler, D. M., Leskovec, J., Mullainathan, S., & Obermeyer, Z. (2021). An algorithmic approach to reducing unexplained pain disparities in underserved populations. Nature Medicine, 27(1), 136-140.

567 Moseley, A., & Thomann, E. (2021). A behavioural model of heuristics and biases in frontline policy implementation. Policy & Politics, 49(1), 49-67.

568 Filmer, D., Nahata, V., & Sabarwal, S. (2021). Preparation, Practice, and Beliefs. World Bank Policy Research Paper 9847.

569 Neary, C., Naheed, S., McLernon, D. J., & Black, M. (2021). Predicting risk of postpartum haemorrhage: a systematic review. BJOG: An International Journal of Obstetrics & Gynaecology, 128(1), 46-53. Khan, M. S., Bates, D., & Kovacheva, V. P. (2021). The Quest for Equitable Health Care: The Potential for Artificial Intelligence. NEJM Catalyst Innovations in Care Delivery, 2(6).

570 Venkatesh, K. K., et al. (2020). Machine learning and statistical models to predict postpartum hemorrhage. Obstetrics and Gynecology, 135(4), 935.

571 https://towardsdatascience.com/the-ultimate-guide-to-starting-ai-d506255d7ea

572 Rahwan, I., et al. (2019). Machine Behaviour. Nature, 568 (7753), 477-486.

573 https://medium.com/behavior-design-hub/what-is-behavioral-data-science-and-how-to-get-into-it-e389ed20751f; https://www.researchgate.net/publication/349963864_Anthropomorphic_Learning

574 Ruggeri, K., Benzerga, A., Verra, S., & Folke, T. (2020). A behavioral approach to personalizing public health. Behavioural Public Policy, 1-13.

575 O'Shea, L. (2019). Future Histories: What Ada Lovelace, Tom Paine, and the Paris Commune Can Teach Us About Digital Technology. Verso Books.

576 https://thedecisionlab.com/insights/technology/this-is-personal-the-dos-and-donts-of-personalization-in-tech

577 Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. The American economic review, 728-741.

578 Camerer, C. F. (2018). Artificial intelligence and behavioral economics. In The economics of artificial intelligence: An agenda (pp. 587-608). University of Chicago Press.

579 Lorenz-Spreen, P., Geers, M., Pachur, T., Hertwig, R., Lewandowsky, S., & Herzog, S. M. (2021). Boosting people's ability to detect microtargeted advertising. Scientific Reports, 11(1), 1-9.

580 See the discussion of "the gaze" in cultural theory. Sartre, J. (1943), Being and Nothingness. Foucault, M. (1975). Discipline and Punish: The Birth of the Prison. Mulvey, L. (1989). Visual pleasure and narrative cinema. In: Visual and other pleasures (pp. 14-26). Palgrave Macmillan, London. Nielsen, C. R. (2011). Resistance through re-narration: Fanon on de-constructing racialized subjectivities. African Identities, 9(4), 363-385.

581 Lewis Jr, N. A. (2021). What counts as good science? How the battle for methodological legitimacy affects public psychology. American Psychologist, 76(8), 1323.

582 Milner IV, H. R. (2007). Race, culture, and researcher positionality: Working through dangers seen, unseen, and unforeseen. Educational Researcher, 36(7), 388-400.

583 Nagel, T. (1986) The View from Nowhere. Robert Sugden uses this term to criticize behavioral economists because, he argues, they implicitly take this stance to make recommendations for society. But this is slightly different from my point, which is not aimed at rejecting paternalism as such, but instead is concerned with the situated nature of any activity, including just observation. Sugden, R. (2013). The behavioural economist and the social planner: to whom should behavioural welfare economics be addressed? Inquiry, 56(5), 519-538.

584 Liscow, Z., & Markovits, D. (2022). Democratizing Behavioral Economics. Yale Journal on Regulation, 39(1217).

585 Bergman, P., Lasky-Fink, J., & Rogers, T. (2020). Simplification and defaults affect adoption and impact of technology, but decision makers do not realize it. Organizational Behavior and Human Decision Processes, 158, 66-79. Pereira, M. M. (2021). Understanding and reducing biases in elite beliefs about the electorate. American Political Science Review, 115(4), 1308-1324.

586 Furnas, A. C (2022) The People Think What I Think: False Consensus and Elite Misperception of Public Opinion.

587 This is not always the case - see Broockman, D. E., & Skovron, C. (2018). Bias in perceptions of public opinion among political elites. American Political Science Review, 112(3), 542-563.

588 Behavioral Insights Team (2018). Behavioural Government, p.35.

589 https://chicagobeyond.org/researchequity/

590 Lewis Jr, N. A. (2021). What counts as good science? How the battle for methodological legitimacy affects public psychology. American Psychologist, 76(8), 1323.

591 https://gocommonthread.com/work/global-gavi-bi/

592 Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. Trends in Cognitive Sciences. https://doi.org/10.1016/j.tics.2022.09.015

593 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5293198/; https://www.apa.org/workforce/data-tools/demographics; Callahan, J. L., Smotherman, J. M., Dziurzynski, K. E., Love, P. K., Kilmer, E. D., Niemann, Y. F., & Ruggero, C. J. (2018). Diversity in the professional psychology training-to-workforce pipeline: Results from doctoral psychology student population data. Training and Education in Professional Psychology, 12(4), 273.

594 Lepenies, R., & Małecka, M. (2019). Behaviour change: extralegal, apolitical, scientistic? In Handbook of behavioural change and public policy (pp. 344-360). Edward Elgar Publishing.

595 Lepenies, R., & Małecka, M. (2019). https://www.psychologicalscience.org/observer/grand-challenges

596 Lewis Jr, N. A. (2021). What counts as good science? How the battle for methodological legitimacy affects public psychology. American Psychologist, 76(8), 1323.

597 Saini, A. (2020). Want to do better science? Admit you're not objective. Nature, 579(7798), 175-176. Dupree, C. H., & Kraus, M. W. (2022). Psychological science is not race neutral. Perspectives on Psychological Science, 17(1), 270-275.

598 Cheon, B. K., Melani, I., & Hong, Y. Y. (2020). How USA-centric is psychology? An archival study of implicit assumptions of generalizability of findings to human nature based on origins of study samples. Social Psychological and Personality Science, 11(7), 928-937.

599 Adetula, A., et. al. (2022) Psychology should generalize from — not just to — Africa. Nat Rev Psychol (2022). https://doi.org/10.1038/s44159-022-00070-y

600 Adorno, T. & Horkenheimer, M. (1972) Dialectic of Enlightenment. Harding, S. (1998). Is science multicultural: Postcolonialisms, feminisms, and epistemologies.

601 Lepenies, R., & Małecka, M. (2019).

602 https://www.urban.org/urban-wire/equitable-research-requires-questioning-status-quo

603 Sendhil Mullainathan, keynote address to the Society of Judgment and Decision Making Annual Conference, 2022.

604 Amin, A. B., Bednarczyk, R. A., Ray, C. E., Melchiori, K. J., Graham, J., Huntsinger, J. R., & Omer, S. B. (2017). Association of moral values with vaccine hesitancy. Nature Human Behaviour, 1(12), 873-880.

605 https://chicagobeyond.org/researchequity/

606 Eriksen, T. H. (2001). Small places, large issues: An introduction to social and cultural anthropology.

607 https://chicagobeyond.org/researchequity/

608 Pereira, M. M. (2021). Understanding and reducing biases in elite beliefs about the electorate. American Political Science Review, 115(4), 1308-1324.

609 https://www.ideas42.org/blog/street-smarts-in-the-field-insights-for-designing-public-housing-infrastructure/

610 Sutherland, R. (2019). Alchemy: The surprising power of ideas that don't make sense. Random House.

611 Sulik, J., Bahrami, B., & Deroy, O. (2021). The Diversity Gap: when diversity matters for knowledge. Perspectives on Psychological Science, 17456916211006070.

612 https://www.poverty-action.org/blog/locally-grounded-research-strengthening-partnerships-advance-science-and-impact-development

613 https://phdproject.org/

614 Erosheva, E. A., Grant, S., Chen, M. C., Lindner, M. D., Nakamura, R. K., & Lee, C. J. (2020). NIH peer review: Criterion scores completely account for racial disparities in overall impact scores. Science Advances, 6(23), eaaz4868.

615 Ledgerwood, A., et al. (2022). The pandemic as a portal: Reimagining psychological science as truly open and inclusive. Perspectives on Psychological Science, 17456916211036654.

The Behavioural Insights Team uses behavioral science to develop better systems, policies, products and services. Our goal is to help people, communities and organizations thrive.

THE
BEHAVIOURAL
INSIGHTS
TEAM

Michael Hallsworth